Driver conflict response during supervised automation: Do hands on wheel matter?

Citation for the original published paper (version of record):

Pipkorn, L., Victor, T., Dozza, M. et al (2021). Driver conflict response during supervised automation: Do hands on wheel matter?. Transportation Research Part F: Traffic Psychology and Behaviour, 76: 14-25. http://dx.doi.org/10.1016/j.trf.2020.10.001

Transportation Research Part F 76 (2021) 14–25

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Driver conflict response during supervised automation: Do hands on wheel matter?

Linda Pipkorn [a,*], Trent W. Victor [a,b], Marco Dozza [a], Emma Tivesten [b]

[a] Chalmers University of Technology, Göteborg, Sweden
[b] Volvo Cars Safety Centre, Volvo Cars, Göteborg, Sweden

ARTICLE INFO

Article history:
Received 23 December 2019
Received in revised form 14 August 2020
Accepted 4 October 2020
Available online 24 November 2020

Keywords:
Automated driving
Human-automation interaction
Wizard-of-Oz
Trust in automation
Driver behaviour
Response process

ABSTRACT

Securing appropriate driver responses to conflicts is essential in automation that is not perfect (because the driver is needed as a fall-back for system limitations and failures). However, this is recognized as a major challenge in the human factors literature. Moreover, in-depth knowledge is lacking regarding mechanisms affecting the driver response process. The first aim of this study was to investigate how driver conflict response while using highly reliable (but not perfect) supervised automation differ for drivers that (a) crash or avoid a conflict object and (b) report high trust or low trust in automation to avoid the conflict object. The second aim was to understand the influence on the driver conflict response of two specific factors: a hands-on-wheel requirement (with vs. without), and the conflict object type (garbage bag vs. stationary vehicle). Seventy-six participants drove with highly reliable but supervised automation for 30 min on a test track. Thereafter they needed to avoid a static object that was revealed by a lead-vehicle cut-out. The driver conflict response was assessed through the *response process*: timepoints for driver surprise reaction, hands-on-wheel, driver steering, and driver braking. Crashers generally responded later in all actions of the response process compared to non-crashers. In fact, some crashers collided with the conflict object without even putting their hands on the wheel. Driver conflict response was independent of the hands-on-wheel requirement. High-trust drivers generally responded later than the low-trust drivers or not at all, and only high trust drivers crashed. The larger stationary vehicle triggered an earlier surprise reaction compared to the garbage bag, while hands-on-wheel and steering response were similar for the two conflict object types. To conclude, crashing is associated with a delay in all actions of the response process. In addition, driver conflict response does not change with a hands-on-wheel requirement but changes with trust-level and conflict object type. Simply holding the hands on the wheel is not sufficient to prevent collisions or elicit earlier responses. High trust in automation is associated with late response and crashing, whereas low trust is associated with appropriate driver response. A larger conflict object trigger earlier surprise reactions.

* Corresponding author at: Chalmers University of Technology, Department of Mechanics and Maritime Sciences, Division of Vehicle Safety, Hus Saga, Hörselgången 4, floor 4, 412 96 Gothenburg, Sweden.

E-mail address: linda.pipkorn@chalmers.se (L. Pipkorn).

https://doi.org/10.1016/j.trf.2020.10.001
1369-8478/© 2020 The Authors. Published by Elsevier Ltd.

1. Introduction

Safe vehicle automation has great potential to save lives by overcoming human limitations, as 94% of crashes are attributed to driver-related critical reasons (NHTSA, 2015). On the other hand, introducing automation to a human–machine system comes with serious human factors challenges (Bainbridge, 1983; Lee, Wickens, Liu, & Boyle, 2017; Seppelt & Victor, 2016) which may jeopardize safety. Today's vehicles already offer driver assistance (*supervised* automation), for which control is shared between the vehicle and the driver at all times (Seppelt & Victor, 2016). That is, the vehicle assists the engaged and attentive driver with longitudinal and lateral vehicle control. Driver engagement and attention is needed because of known system limitations (e.g., sensors that may not detect a pedestrian crossing a highway or an object in lane) that might occur unexpectedly and that require driver intervention.

How an extended period of supervised automation affects driver response to automation limitation is today not well understood (McDonald et al., 2019; Mole et al., 2019). Current evidence on how drivers respond in critical situations after automation mainly stems from studies of *unsupervised* automation (McDonald et al., 2019; Mole et al., 2019). This type of automation does not require the driver to be attentive or engaged since the automated driving system (ADS) prompts the driver response by alerting the driver to take back control (the transition of control is *ADS-prompted*) through a salient notification when needed. In contrast, in supervised automation the driver response can either be *ADS-unprompted* when there is no notification given by the system, but the driver detects the need to act when system limitations occur (e.g., an unexpected object in lane which the system does not detect, or a steering torque limitation in a curve), or ADS-prompted (e.g., a disengagement message, or a forward collision warning). *System limitations* (i.e., a declared limitation of the system, often described in the car manual) should be differentiated from a *silent-failure* event where the system is not working as intended. There is currently a lack of understanding of how drivers respond in ADS-unprompted *system limitation events*, even if the literature on driver response to *silent-failure events* is expanding (Bianchi Piccinini et al., 2019; Larsson, Kircher, & Hultgren, 2014; Strand, Nilsson, Karlsson, & Nilsson, 2014).

The change of tasks and situation for the driver, as a consequence of reliable supervised and unsupervised automation in contrast to manual driving, is associated with (a) reduced situation awareness or entering an "out-of-the-loop" state (Endsley & Kiris, 1995) and (b) degraded performance when automation fails after having performed well for a long period (Bainbridge, 1983; Wickens, Hooey, Gore, Sebok, & Koenicke, 2009). The out-of-the-loop driver state is defined by Merat et al. (2019) as: "Not in physical control of the vehicle, and not monitoring the driving situation, OR in physical control of the vehicle but not monitoring the driving situation", in contrast to the in-the-loop state: "In physical control of the vehicle and monitoring the driving situation", and the on-the-loop state, "Not in physical control of the vehicle, but monitoring the driving situation". However, the details of what constitutes physical control and monitoring is not fully understood (Mole et al., 2019; Victor et al., 2018). Merat et al. (2019) offers a first definition of what physical control means in terms of driver engagement: "Vehicle physical control implies a direct physical coupling between multisensory perceptual cues and motor outputs (steering, accelerating/braking". In addition, Merat et al. (2019) points out that: "being in-, on– and out of the loop should not be viewed as discrete states, but rather levels of engagement along a continuum.". Thus, keeping the hands on the wheel is the first sign of the driver being in physical control and in-the-loop, but to what extent the driver also needs to apply torque or frequent steering input is still unknown. Investigations of recent crashes involving vehicles with supervised automation (NTSB, 2017, 2018, 2019) have found longer durations of hands-off-wheel driving, which indicates that drivers with a reduced physical control is already present in real traffic. According to the Merat et al. (2019) definition, these drivers may be either out-of-the-loop or on-the-loop dependent on their monitoring behaviour. Hands-off-wheel driving is regulated differently in the US and the EU: in the US, most states do not specifically prohibit hands-off-wheel driving (Smith, 2013), while current European Regulations do not allow hands-off-wheel driving for longer than 15–30 s without a warning (visual at 15 s and visual and auditory at 30 s) (UNECE, 2017). To be able to inform current hands on wheel regulations which today differ between the US and the EU it is important to understand if requiring drivers to keep their hands on the wheel or not influences driver conflict response.

As a step toward better understanding the driver conflict response to ADS-unprompted system limitation events, Victor et al. (2018) investigated conflict outcome in a longitudinal cut-out scenario (Euro NCAP, 2018) taking place after 30 min of supervised automation. In their study, one-third of the drivers (all high-trust) crashed with the stationary conflict object, which was revealed by the lead-vehicle cut-out. The drivers crashed because they expected automation to act, but automation could not detect the object (because it was a declared system limitation) and therefore did not act (Victor et al., 2018). The crash rates were similar with or without a hands-on-wheel requirement, and for the type of conflict object (e.g., stationary vehicle, garbage bag). Additional analysis revealed that almost all crashers (95%) reported (after the drive) that they expected automation to act, despite being informed otherwise (Gustavsson et al., 2018). In addition, 77% of the crashers either reported that they did not realize the need to act or realized the need too late. On the other hand, most non-crashers (74%), reported that they did expect or were uncertain about an intervention from automation and that they realized the need to act or realized the need late (85% non-crashers). Further analyses by Tivesten, Victor, Gustavsson, Johansson, and Ljung Aust (2019) investigated glance behaviour during the 30 min of driving prior to the conflict and found that three different pre-conflict behavioural patterns were indicative of increased risk of crash involvement. These patterns were: low levels of visual attention to the forward path, high Percent Road Centre (i.e., gaze concentration), and long visual response times to attention reminders. Out of these patterns, low visual attention was the most common behaviour present in 70% of the crashers. To conclude, Victor et al. (2018) found high trust in automation to be associated with crashing, whereas the conflict object type did not seem to influence if drivers crashed or not. However, a detailed understanding of how (i.e., in which way) these factors influences driver conflict response in automation is missing. Specifically, the influence of trust in automation on driver conflict response has been pointed out as a factor that needs additional work (McDonald et al., 2019). The influence of different conflict object types on the driver conflict response in automation has not received much attention previously (Mole et al., 2019; McDonald et al., 2019), but is important to consider since it may impact driver conflict response.

The present paper expands upon this previous work (Gustavsson et al., 2018; Tivesten et al., 2019; Victor et al., 2018) by analysing the driver conflict response in ADS-unprompted conflicts due to system limitations. A more detailed understanding of the driver conflict response is provided through the *response process*: timepoints for driver actions prior to either crashing or avoiding the conflict object. The first aim of the present study is to understand how the conflict response differ between drivers that (a) crashed and drivers that avoided the conflict object and (b) reported high trust in automation to handle the conflict and drivers that reported low trust in automation to handle the conflict. The second aim was to understand the influence of two specific factors: hands-on-wheel requirement (with vs. without) and the conflict object type (garbage bag vs. stationary vehicle), on driver conflict response. Specifically, the following research questions were considered: (1) Is the conflict outcome (crash vs. no crash) associated with a specific type of response process?, (2) Is the subjectively reported level of trust in automation to handle the conflict (high vs. low) associated with a specific type of response process?, and (3–4) Is the response process influenced by: a hands-on-wheel requirement (with vs. without), and the type of conflict object (garbage bag vs. stationary vehicle)?

2. Methods

The data was collected in two test track experiments previously reported (study 2 and 3 in Victor et al., 2018). Seventy-six participants drove with supervised automation (30 min), before experiencing a longitudinal cut-out scenario (Euro NCAP, 2018) with a stationary static conflict object (CO) in lane, requiring the participants to intervene to avoid a crash. The independent variables were: (a) the hands-on-wheel (HoW) variable (30 participants were required to supervise automation with hands on the steering wheel, and the remaining 46 (16 in study 2, 30 in study 3; see Table 1) participants were not required to) and (b) the CO variable (30 participants encountered a stationary vehicle and 46 (16 in study 2, 30 in study 3; see Table 1) participants encountered a garbage bag as the conflict object).

2.1. Equipment

The experiments were performed on the AstaZero test track (rural road) near Gothenburg, Sweden. The Test Vehicle (TV) was a Volvo XC90 (Model Year 2016) equipped with standard sensors, a GPS, and three cameras (collecting data at 20–30 Hz) and with a modified driver information display (i.e., to present experiment-specific attention reminders). The TV used a *Wizard-of-Oz* method (Habibovic, Andersson, Nilsson, Lundgren, & Nilsson, 2016) with a safety driver in the back seat. In addition, a test leader, collecting subjective measurements throughout the drive, was also seated in the back seat. A robot-controlled XC90 lead vehicle (LV) was the only other vehicle present.

2.2. Scenario design

All participants supervised automation for 30 min before experiencing the conflict situation of interest for this paper. Before the test, the participants received information about the automated system limitations and the supervising role. As described in Victor et al. (2018), all 16 participants in study 2 received medium detailed written instructions (all in the without HoW requirement and Bag group in Table 1). All 60 participants in experiment 3 received high detailed classroom training (and were equally distributed with 15 participants per group in Table 1), see Victor et al. (2018) for further details. The participants were informed to continuously supervise the drive (as they would in normal driving) and to intervene whenever needed. Further, the automated system's limitations were emphasized and examples of situations requiring driver intervention were explained (e.g., potholes, objects in lane). In addition, the participants were informed about the possibility to override automation (i.e., steer or brake) at any time. Throughout the drive, the TV followed the LV at a constant headway and with a maximum speed of 70 km/h. No conversation between the participant, the test leader and the safety

Table 1
Group Specifications including number and percentage of participants.

Participants [N (%)]	Without HoW requirement		With HoW requirement		Total
	Car	Bag	Car	Bag	
Study 2	0	16	0	0	
Study 3	15	15	15	15	
Total	***15 (20%)***	***31 (40%)***	***15 (20%)***	***15 (20%)***	***76***

driver took place, except when asking for subjective measures. All participants received attention reminders if they had repeated or long glances away from the road (Victor et al., 2018). The attention reminders were triggered by the test leader based on a specific glance behavior algorithm. In practice, the test leader observed the driver from a video in the backseat and manually applied the algorithm by detecting and measuring eyes off path durations. The attention reminders were given at different levels: the first level attention reminders were given as visual warning messages (‘‘Stay attentive! Look ahead" in study 2, ‘‘Driver inattention – Keep eyes on road" in study 3) in the instrument cluster. If the first level reminder was ignored, the drivers could obtain higher level reminders (level 2 in study 2 and level 2–3 in study 3). Higher level reminders included additional information (e.g., audio, red icon) to make sure the driver would receive them.

2.2.1. The hands-on-wheel requirement and reminders

The seventy-six participants were divided into two groups: 30 participants were required to supervise automation with their hands on the wheel (with HoW requirement) and 46 were not required to (without HoW requirement). In addition to the attention reminders, the participants with HoW received reminders to keep their hands on the wheel if they did not. In accordance with the attention reminders, the HoW reminders were triggered by the test leader if the driver took their hands off the wheel for more than 5 s (level 1), or more than 10 s (level 2). The first level constituted of a visual message (‘‘Driver inattention – Apply steering") which was displayed in the instrument cluster, and the second level constituted of the same visual message combined with a sound.

2.2.2. The conflict situation

The longitudinal cut-out scenario, used as the conflict situation and which appeared after 30 min of driving, included two different conflict object (CO) types: a garbage bag or a stationary ‘‘balloon" vehicle. The CO was visible for the first time when passing through a curve and over a crest, approximately 11 s (206 m) before reaching the CO. The CO became visually obstructed again (TV about 8 s from CO) by the LV when the road straightened out and appeared again when the LV performed a sudden cut-out (TV about 3 s or 60 m from the CO). The TV did not brake or warn the driver in any way (see Fig. 1), and the driver had to override automation (i.e., steer, brake) to avoid a crash.

After the conflict, each participant reported their level of trust in automation to handle the conflict with response on a 7-point scale (1 = not at all, and 7 = completely).

2.3. The driver response process

To assess driver response during conflicts, a driver response process was defined as timepoints for a surprise reaction (SRT), hands-on-wheel (HOW), driver steering (DS) and driver braking (DB) taking place prior to either avoiding or crashing with the conflict object (i.e., the *conflict anchor* t = 0). The response process relevant timepoints can be seen in Table 2.

2.4. Data processing and coding

The timepoints for the response process were coded manually from video views showing the forward path and the drivers face and upper body side. The video views were also used to assess conflict outcome (i.e., if the drivers crashed with or passed the CO). The event was coded as a crash if any contact between TV and CO was present (timepoint t_{crash} at contact onset) and otherwise as pass (timepoint t_{pass} when the longitudinal distance between the TV front and CO reached zero). In addition, three *reference timepoints* were added to establish when certain events took place: (1) when the CO was visible for the first time (see event at 11 s in Fig. 1), (2) the start of the LV cut-out maneuver (see event at 3.5 s in Fig. 1), and (3) when the CO became fully visible (see event at 2.5 s in Fig. 1). To control for the possible influence of off-path glances on the

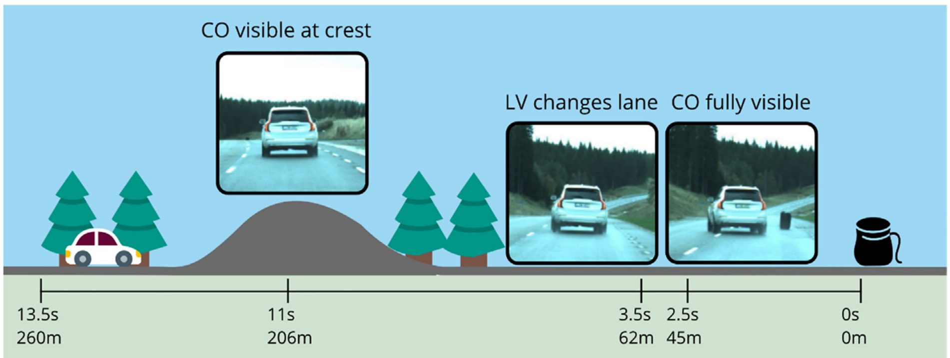

Fig. 1. After 30 min of driving the participants reached a crest from which the CO could be observed the first time. Then, LV covered CO again for some time before cutting out of lane to fully reveal CO when the TV was about 3 s from it.

Table 2
Response Process Relevant Timepoints.

Surprise Reaction Timepoint (SRT):	The time of the first video frame when the driver appears to recognize an unexpected event. This can be a change in facial expression, gaze fixation or movement of a body part in a way that indicates awareness and/or the start of an evasive maneuver. *Example:* The participants supervises the automated driving with the hands placed on the lap. When the CO becomes visible to the participant he/she reacts by extending the neck and adjusting the seating position to better see what is happening further down the road (SRT).
Hands on Wheel Timepoint (HOW):	The first video frame when the driver is seen to touch the steering wheel with at least part of one hand. This is not applicable if one hand is already on steering wheel. *Example:* When the CO becomes visible to the participant he/she might place the hands on the steering wheel (HOW) to be prepared to intervene.
Driver Steering Timepoint (DS):	The first video frame when the driver starts performing a steering maneuver. A binary vehicle signal measuring applied torque on the steering wheel was used to support the decision of the timepoint. *Example:* When the LV changes lane and the CO is revealed the participant starts steering (DS) to avoid crashing into CO.
Brake Timepoint (DB):	This was determined by a binary vehicle signal indicating when the brake pedal was pressed. *Example:* When the LV changes lane and the CO is revealed the participant starts braking (DB) to avoid crashing into CO.
On-path glance	An on-path glance was defined as any glance out of the forward windshield in the direction of travel, including fixations on CO.
Off-path glance and transitions	An off-path glance was defined as all glances that were not defined as an on-path glance, including eye closures longer than 150 ms. Note that the assigned area of interest also included video frames capturing transitions to that area of interest.

response process, the glance behavior during the conflict was also assessed. The on- and off-path glances (see definition in Table 1) were coded using the video views of the forward path and the driver's face.

2.5. Data analysis and visualizations

To answer the research questions, groups for comparisons were formed using information about: (1) the conflict outcome, (2) the experimental conditions for HoW requirement and CO type, and (3) the a posteriori reported trust level.

2.5.1. Analysis 1: The response process and glance behaviour for crashers and non-crashers

This analysis was performed to accommodate the first aim (i.e., to understand how the drivers who crashed and avoided the conflict object responded). The response process was visualized with cumulative distributions for the four variables (SRT, HOW, DS, DB) in a time interval starting at −13.5 s and leading up to the conflict anchor (t_{crash} for crashers and t_{pass} for non-crashers). In case of multiple HOW timepoints the last was included in the response process. In the case of no observable SRT, HOW or DS the time was set to t = 0. Within the same time interval (from −13.5 s to 0 s), the *Percent Road center* (PRC) was calculated at each time step by calculating the percentage participants having eyes on-path. The average and standard deviation PRC were then calculated for seven intervals ($-13.5\text{ s} \leq t < -12\text{ s}$, $-12\text{ s} \leq t < -10\text{ s}$, ..., $-2\text{ s} \leq t < 0$). Included in the same graph as the comparison of crashers and non-crashers (list item 1) was also a comparison of PRC and three vertical lines representing the average value of the three reference timepoints.

2.5.2. Analysis 2–4: The response process for difference in hands-on-wheel requirement, the conflict object type and the trust level

These analyses were performed to accommodate the first and second aim (i.e., to understand how the drivers with high and low trust responded and the influence of a HoW requirement and CO type on the driver conflict response). Three main analyses (numbered 1, 2 and 3 in the list below) were performed. Further, for trust level two additional analyses (2.1, 2.2) were performed to better understand the relation between trust level and conflict outcome. The response process was compared for the following groups:

1. Only study 3 participants: with HoW requirement (n = 30), without HoW (n = 30)
2. All participants except mid-trust drivers: high-trust drivers (n = 46), and low-trust (n = 26)
   2.1. Non-crashers except mid-trust drivers: high trust (n = 25), and low trust (n = 26)
   2.2. High trust participants: crashers (n = 21) and non-crashers (n = 25)
3. Only study 3 participants: conflict object: Bag (n = 30) and Car (n = 30)

Table 3 (an extended version of table 1) displays how the participants (divided into crashers and non-crashers in the top panel and according to trust level in the bottom panel) were distributed for the two independent variables (HoW and CO).

2.5.3. Statistical analysis

To verify statistically significant differences between the groups, a Mann-Whitney *U* test was used for all comparisons for the response times SRT, HOW and DS. A non-parametric test was used because of the data distributions for SRT, HOW and DS

Table 3
Group Specifications.

Participants [N/N (%)]	Without HoW requirement		With HoW requirement		Total
	Car	Bag	Car	Bag	
Crashers	3	9	5	4	21
Non-crashers	12	22	10	11	55
Total	***15***	***31***	***15***	***15***	***76***
High trust	10	16	11	9	46
Mid trust	0	3	0	1	4
Low trust	5	12	4	5	26
Total	***15 (20%)***	***31 (40%)***	***15 (20%)***	***15 (20%)***	***76 (100%)***

were observed (from histograms) to be non-normal. The results were considered statistically significant at 0.01 after a Bonferroni correction for multiple testing (0.05/5), based on the assumption of SRT, HOW and DS being related dependent variables.

3. Results

3.1. Response process in crashers and non-crashers

Fig. 2 shows the PRC and the response process for the (n = 21) crashers and the (n = 55) non-crashers. The PRC is similar for the time intervals after the object was visible for the first time (–11 s, at the crest), while the mean PRC was noticeably lower for crashers (~70%) than for non-crashers (~86%) in the ($-13.5\ s \le t < -12\ s$) interval. Fig. 2 also indicates that all except one participant constantly had the eyes on-road during the last 2 s before the conflict anchor.

Further, Fig. 2 shows that crashing participants (dotted lines) generally showed a surprise reaction, put hands-on-wheel and started steering closer to the conflict object compared to the non-crashing participants (solid lines). In addition, some crashers never put hands-on-wheel nor started steering (at 0 s the HOW and DS curves only reached 67% and 52% respectively), whereas all non-crashers did (at –1 s the HOW and the DS curves reaches 100%). None of the crashers and only 11% (6/55) of the non-crashers braked.

3.2. Influence of a hands-on-wheel requirement

Fig. 3 presents the response process for participants with (n = 30) and without (n = 30) the HoW requirement. The HoW requirement resulted in that 28 participants had HOW at –13.5 s (i.e., only 2 participants did not have hands on wheel, even

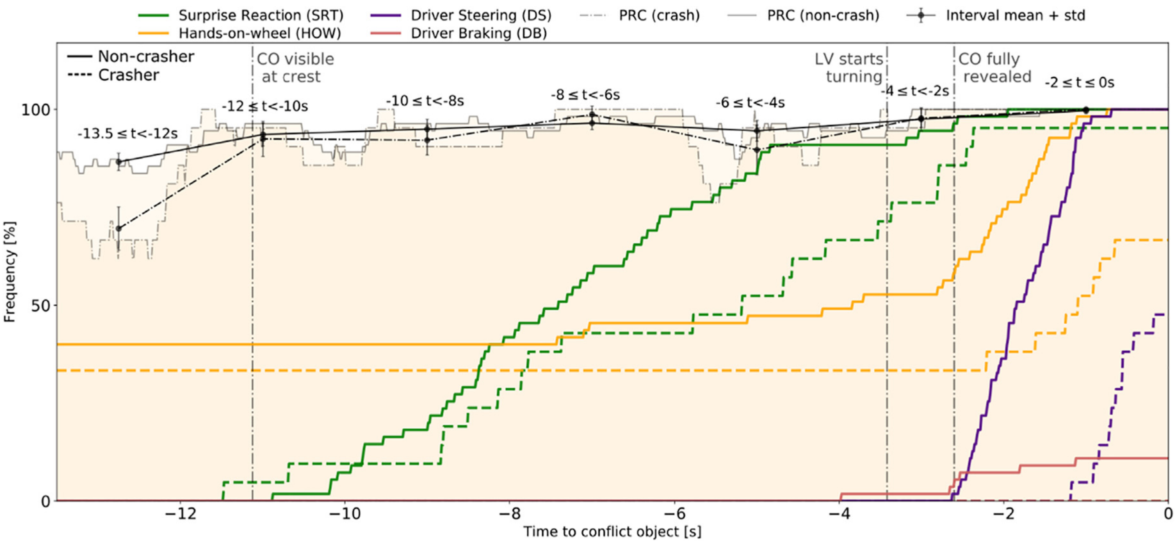

Fig. 2. Response Process and Percent Road Center for crashers (dotted lines) and non-crashers (solid lines). The three vertical lines represent (from left to right) the mean of the time points for when: (1) the CO was visible to the driver the first time, (2) the LV started turning and (3) the CO was fully revealed to the participant.

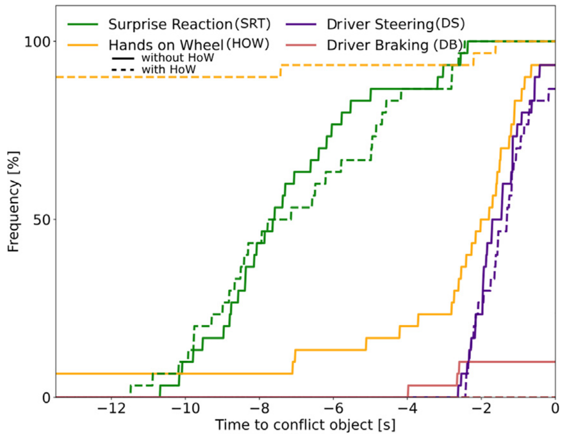

Fig. 3. Response Process for drivers with (dotted lines) and without (solid lines) a HoW requirement.

if required to). But, the HoW requirement did not result in significantly different SRT and DS distributions: the participants without HoW requirement had similar distributions to participants with HoW requirement for SRT (Mdn = –7.62 s, Mdn = –7.46 s) and DS (Mdn = –1.58 s, Mdn = –1.31 s), and did not differ significantly for SRT (U = 437, p = .85, r = –0.02) or DS (U = 408, p = .53, r = –0.08) respectively. None of the participants with the HoW requirement braked, while 10% (3/30) of the participants without HoW requirement did.

3.3. Influence of high and low trust

Fig. 4a shows the response process for the (n = 26) low-trust participants and (n = 46) high-trust participants. The SRT did not differ significantly for the high-trust (Mdn = –7.57 s) and low-trust participants (Mdn = –7.34 s), U = 542, p = .51, r = –0.08. However, for the participants who did not have their hands on wheel from the start, the high-trust participants (Mdn = –1.23 s) put their hands on wheel significantly later than the low-trust participants (Mdn = –2.40 s), U = 90, p = .001, r = –0.49. In addition, the high-trust participants also steered (Mdn = –1.10 s) significantly later than the low-trust participants (Mdn = –1.91 s), U = 239, p < .001, r = –0.48. Braking was more prevalent for low-trust participants (4/26) compared with the high-trust participants (1/46).

Fig. 4b presents the response process for the (n = 26) low-trust and the high-trust (n = 25) non-crashers. The SRT-curves are overlapping, whereas a minor delay can be observed for the high-trust non-crashers compared to the low-trust non-

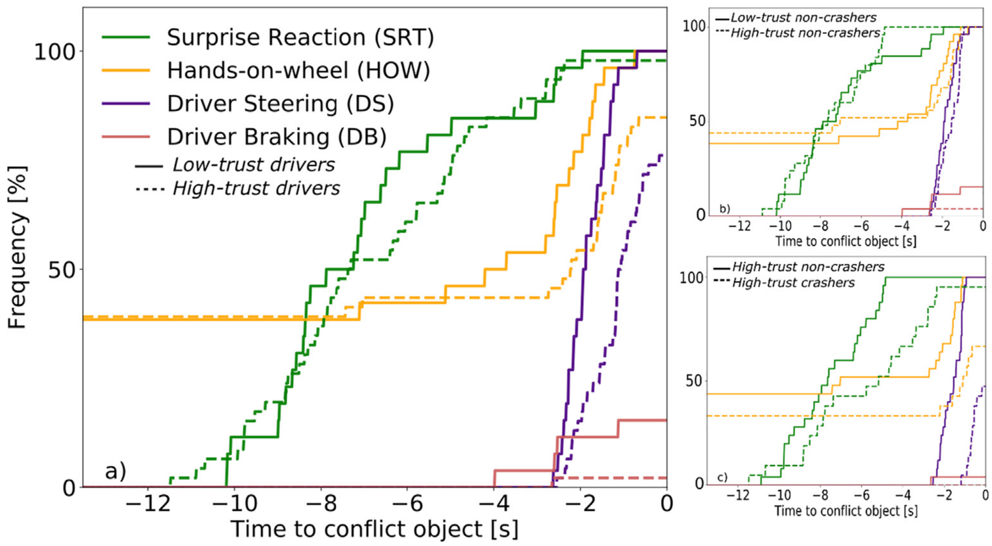

Fig. 4. (a) The response process for all high-trust drivers (dotted lines) and low-trust drivers (solid lines), (b) The response process for high-trust non-crashers (dotted lines) and low-trust non-crashers (solid lines), and (c) The response process for high-trust crashers (dotted lines) and high-trust non-crashers (solid lines).

crashers (the dotted lines are shifted slightly to the right). Fig. 4c presents the response process for the (n = 25) high-trust non-crashers and (n = 21) crashers: The SRT is observed noticeably earlier for the non-crashers compared to the crashers. However, this difference was not statistically significant under the Bonferroni correction, U = 163, p = .028, r = –0.32.

3.4. Influence of conflict object type

The participants in the Bag condition (n = 30) showed a surprise reaction significantly later (Mdn = –6.54 s) than the participants in the Car condition (n = 30) (Mdn = –8.54 s), U = 252, p = .003, r = –0.38, but HOW and DS did not differ significantly across the groups. All participants in the Car condition had hands on wheel at the conflict anchor (30/30), compared to the 93% (28/30) participants in the Bag condition. In addition, 97% (29/30) of the participants in the Car condition steered to avoid the stationary vehicle, compared to 83% (25/30) of the participants in the Bag condition. Braking was slightly more prevalent for the Bag than for the Car (2/30 for Bag, 1/30 for Car).

4. Discussion

4.1. The driver response process and glance behavior for crashers and non-crashers

The present study found that the participants spent most of their time looking on-path during the conflict (i.e., 100% of the participants looked on path at some point during the seconds before the conflict anchor and the PRC was high (>90%) for most of the 13.5 s leading up to the conflict object in Fig. 2). This additional analysis confirms that one-third of the participants crashed, despite having eyes on path, as previously concluded by Victor et al. (2018). That is, the high PRC (>90%) indicates that when the drivers noticed the conflict object (first visible at –11 s) they kept on looking at it, probably to make predictions about what was about to happen and how to respond to the situation. Further, a lower PRC for crashers between –13.5 s and –12 s compared to non-crashers were identified. This result is likely a consequence of that several crashers (70%) were found to have low levels of visual attention to the forward path during the 30 min automated driving prior to the conflict (Tivesten et al., 2019).

The present study extended the previous analysis of conflict outcome reported in Victor et al. (2018), with a detailed investigation of the drivers' response process during the conflict. Generally, crashers were observed to respond later in all actions except braking (i.e., for SRT, HOW, DS) of the response process compared to non-crashers (see Fig. 2). These differences in response for crashers and non-crashers are likely a consequence of some underlying cognitive control mechanism (i.e., the crashers did not understand the need to act, Victor et al., 2018) and is a likely explanation for why some drivers crashed and some did not (i.e., the drivers crashed because they steered late or not at all, compared to the drivers that avoided a crash because they responded earlier).

Also, the present study observed that some crashers did not act at all, whereas some put hands-on-wheel and attempted steering (Fig. 2 indicates that HOW only reached 70% and DS reached only 50% for crashers). Interestingly, the crashers were more likely to act (i.e., put hands-on-wheel and steer) if the conflict object was the stationary vehicle: in section 3.4, 97% of the participants steered in response to the stationary vehicle, but only 83% for the garbage bag. This is likely due to the larger size of stationary vehicle compared to the garbage bag. The larger size of the stationary vehicle is likely also the reason behind the observed earlier surprise reaction in the Car condition compared to the Bag condition (see Section 3.4).

4.2. The influence of requiring drivers to keep hands on the wheel on the driver response process

Victor et al. (2018) concluded that drivers crashed to the same extent independent of if they had their hands on the wheel or not. The present study enriched this understanding further by finding that requiring a participant to keep hands on wheel also did not result in earlier steering response times. Together, these findings indicate that driver conflict response with a static object in supervised automation is independent of a HoW requirement.

The present study found that longer periods of hands-off driving does not necessarily lead to decreased conflict intervention performance – a result in line with previous studies on ADS-prompted conflicts (Naujoks, Purucker, Neukum, Wolter, & Steiger, 2015; Naujoks et al., 2017). Notably, most drivers in these previous studies kept contact with the steering wheel even if allowed to have hands off. On the other hand, Llaneras, Cannon, and Green (2017) indicated that using a HoW requirement as an *escalation of consequence* for drivers who ignore visual attention reminders resulted in more successful driver interventions to a silent failure lane-drift event (i.e., a lateral conflict situation). Although the present study does not find support for improved driver conflict response with a HoW requirement in a longitudinal cut-out scenario, it is possible that other conflict types would benefit from hands on wheel. For example, a driver with hands on wheel may reflexively and beneficially counter-steer in an abrupt incorrect steering (as indicated by Benderius, 2014). Further, a HoW requirement might prevent drivers from attending to visual-manual secondary tasks. In addition, besides the requirement to keep the hands on the wheel, a key parameter in designing safe supervised automation may be the lane keeping performance and characteristics of the automated system. The driver is likely to provide more frequent steering input, and have a lower trust in the system, if the lane keeping performance is less precise (e.g., varying lane position) or less reliable (e.g., leaving lane without driver input). More research on the parameters and thresholds determining how to keep the drivers "in-the-loop" and sufficiently engaged in the driving task (i.e., in physical control) is needed.

4.3. The influence of trust on the driver response process

The present study enriches understanding of the influence of trust on driver conflict response, beyond the previous finding of higher level of trust in crashers compared to non-crashers (Victor et al., 2018; Gustavsson et al., 2018). Firstly, a high-trust participant was found to generally respond later than a low-trust participant – a finding in line with (Körber, Baseler, & Bengler, 2018) who found that high trust results in more collisions. Second, the high-trust non-crashers were found to put hands on wheel and start steering at similar points in time compared to the low-trust participants (all non-crashers; see Fig. 4b). Third, high-trust non-crashers were observed to show a surprise reaction earlier (visible difference in Fig. 4c, even if not statistically different after the Bonferroni correction) than a high-trust crasher. In addition, it should be noted that the different trust levels (i.e., high, low trust) also corresponded to different reported experiences: the majority of high-trust participants reported that they expected automation to act, whereas most low-trust participants did not expect automation to act (Gustavsson et al., 2018). Taken together, these results suggest that high-trust- and low-trust drivers have different mental representations of how the automation would behave in the conflict situation, and this results in later response for high-trust drivers. For high-trust crashers compared to high-trust non-crashers, the delay in response seems to stem from a later surprise reaction.

4.3.1. Applying the predictive processing framework on the results

One promising framework to explain the observed delay in response for high-trust drivers compared to low-trust drivers is the predictive processing (PP) framework (Clark, 2013; Engström et al., 2018). According to the PP framework, the brain continuously predicts sensory input from the external environment (e.g., looming – the visual expansion of an approaching lead vehicle on the retina) and minimizes deviations between predicted and perceived sensory inputs, through action (e.g., braking, steering) or by updating the prediction. These predictions are generated by a hierarchical generative model, which is embodied in the brain and develops with experience. In the context of manual driving, the continuous minimization of prediction errors by either updating predictions or acting (steering, braking) is referred to as active inference. Further, active inference may take place on different hierarchical levels of the driving task (i.e., on the operational, tactical and strategic levels according to Michon, 1985).

We suggest that the delay in the response process or the absence of a response, as observed in the present study, can be explained by differences in the generative models (i.e., different mental representations of automation capabilities) in the PP framework. How different generative models may lead to delayed response, was previously argued in a work on silent adaptive cruise control (ACC) failures by Bianchi Piccinini et al. (2019). Longer brake response times were found for ACC-drivers compared to CC-drivers in a braking lead vehicle scenario, when the ACC-system failed silently. The longer brake response times, for ACC-drivers, were explained to arise from the ACC-drivers' generative model of a working ACC-principle. When the lead vehicle braked, the ACC-drivers' generative model allowed more looming (since they expected ACC to brake) compared to the CC-drivers who braked at lower looming levels. In the context of the present study, the differences in expectations on how automation would act in the conflict situation, suggests that low-trust- and high-trust drivers embody different generative models. The high-trust drivers seem to embody a generative model that predicted automation to act in the conflict situation, whereas low-trust drivers did not predict automation to act. Consequently, at the conflict onset (when the lead vehicle performed the cut-out and looming suddenly increased) the high-trust drivers accepted more looming before acting (since they awaited automation to act), compared to the low-trust drivers.

However, in comparison to Bianchi Piccinini et al. (2019), the present study also observed that having similar generative models (as assumed for the high-trust drivers) does not necessarily lead to the same response. Among the high-trust drivers some responded too late or not at all and crashed, and the rest responded early and did not crash. The reason behind the difference in response among high-trust drivers could be that several other factors in addition to the generative models also influences response (e.g., criticality of the situation, individual response times, when the drivers detect the conflict etc.). Another possible explanation for the difference in response among high-trust drivers is that these drivers were engaged in inference at different levels of the driving task. According to Engström et al. (2018), automation changes the inference from active to *perceptual inference*, on the operational, tactical and strategic levels. The reason is that the driver engages in monitoring the automated system, rather than actively cancelling prediction errors by steering and braking. In supervised automation, Engström et al. (2018) argues, the drivers can still be engaged in perceptual inference on the operational level but may also only be engaged in inference on higher levels (i.e., tactical, strategic). A driver who remains in the operational loop monitors the lead vehicle and generates predictions about the looming, whereas a driver who only engages in inference on higher levels monitors the general situation but does not predict looming.

The explanation for why some high-trust drivers crashed and some did not, may thus be that the non-crashers remained in the operational loop, but the crashers did not. This assumption is supported by the finding of low levels of visual attention to the forward path during the complete drive among the majority of the crashers (Tivesten et al., 2019) and the delayed surprise reaction among crashers. A high-trust driver, who is out of the operational loop (i.e., does not predict looming), will act later because of the need to re-establish predictions at the operational level (i.e., predict looming). When these predictions are re-established, prediction errors can be generated and acted upon (Engström et al., 2018). The reason behind why some crashers did not act at all, may be because they never re-established predictions on an operational level (i.e., no predicted looming means no prediction errors to act upon). On the other hand, the crashers that did act, may have re-established predictions on the operational level, but too late to accumulate enough evidence for action (prediction error). Another potential explanation is that all crashers re-established predictions at similar times, but the larger conflict object triggered earlier response for the crashers that acted (larger looming means faster accumulation of evidence for action). In fact, it could be that drivers who are engaged in inference at different levels of the driving task embody different generative models (i.e., that the high-trust drivers that crash and the ones that avoided the conflict had different generative models). If this is true, the simple division of drivers into high-trust- and low-trust drivers performed in this study does not seem to be able to capture those differences.

Finally, another influencing factor behind the delayed response observed for high-trust drivers, may be a lower *gain* (Bianchi Piccinini et al., 2019). That is, these drivers accumulate looming slower, and consequently it will take longer to reach the prediction error that triggers an action. The reason behind the lower gain may be due to individual factors or the result of exposure to automation, but a model based on the lower gain hypothesis cannot distinguish between these.

Future work should study the underlying cognitive mechanisms ruling driver response process in conflict situations by systematically manipulating different factors that may influence the driver conflict response e.g., conflict expectancy, different degrees of driver awareness of system limitations. For example, would adding a prompt (e.g., a take-over request in unsupervised automation, a warning in supervised automation) in the same situation override potential problems such as slow accumulation of evidence and/or late surprise reactions, and therefore prevent crashes? Based on the PP framework, adding a prompt to the longitudinal cut-out scenario is assumed to result in earlier conflict response, because it triggers the drivers to start making predictions about looming at an earlier stage. Consequently, the earlier the prompt is given, the earlier the drivers can be assumed to start to make predictions about looming and therefore calculating prediction error and act upon it.

4.4. Limitations and practical applications

The results presented in this paper should be viewed in the light of its limitations. The experiment was performed on a test track and not in real traffic. Thus, there was some lack of realism as no traffic was present, the conflict object was a balloon vehicle or a garbage bag, the participants knew that they were part of a study and a test leader was present. However, it is difficult—if not impossible—to increase the degree of realism without encountering difficult ethical considerations associated with the risk of crashing. In addition, video coding was performed by one single person. A further limitation is that the drivers answered the questionnaires after the study and therefore the (crash/no-crash) outcome of the study might have influenced how they responded. Specifically, the results on the influence of driver trust level on the driver conflict response should be treated as associative rather than casual. Recall that the same is true for the analyses including the conflict outcome. In addition, it should be mentioned that the present study only considered the timing of the driver conflict response (when the drivers started steering) and not the quality (the steering performance). Instead, the driver conflict response quality in the present paper is only assessed in terms of the drivers' ability to avoid crashing with the object. For studies where there is no clear understanding of how a late response may be linked to the conflict outcome, it is important to consider both the timing and the quality of the driver conflict response. The reason is that the timing of the driver conflict response does not necessarily correlate with the quality of the response (Louw et al., 2017).

The present study has some practical implications: we did not find any evidence for faster response for drivers that were required to keep their hands on the wheel. This result can inform current hands on wheel regulations (at least for the cut-out conflict situation): a hands-on-wheel requirement alone is not enough to prevent collisions or prompt earlier responses in highly reliable (but not perfect) supervised automation. Note that the findings from the present study apply to a longitudinal cut-out scenario with a stationary object, the same results might not apply for lateral conflicts (e.g., lane drifts, or steering failures). In addition, the results in the present study suggests that supervised automation should be designed in a way that facilitates low automation trust, since low trust is associated with appropriate driver conflict response. On the other hand, it is important not to forget that too low automation trust may also result in *disuse* (Parasuraman & Riley, 1997), i.e., too low trust may result in drivers that do not use the system at all. Thus, supervised automation should be designed in a way that facilitates low enough trust to ensure appropriate conflict response, but high enough trust for the drivers to utilize it (i.e., not turn it off because they do not trust it).

5. Conclusions

Previous studies found that, some drivers, after having supervised automation for some period, may crash with a suddenly appearing conflict object that is not detected by automation, even if this is a declared system limitation. By performing detailed analysis of the driver response process, the present study also found that these drivers (i.e., the crashers) generally show a late surprise reaction and put hands on wheel, and start steering closer to the conflict object (i.e., a delayed response) compared to the drivers that do not crash (i.e., the non-crashers). In fact, some drivers crashed without putting their hands on the wheel nor steering or braking.

Further, a requirement to just keep the hands-on-wheel may not prevent drivers from crashing, nor promote faster response. A certain minimum level of applied torque to the steering wheel is likely needed to keep all drivers in the loop during extended periods of supervised automation. The extent to which a hands-on-wheel requirement may still be beneficial and prevent other types of conflicts (e.g., lateral automation failures, incorrect steering) or keep the driver from performing visual-manual secondary tasks needs further investigation. High trust is associated with late response and crashing. However, high trust in combination with early surprise reaction is associated with crash avoidance. Low trust in automation, on the other hand, is associated with appropriate driver conflict response. Finally, a larger conflict object may trigger an earlier surprise reaction but does not hasten hands-on-wheel or steering response.

To conclude, we found that looking at the response process together with the glance behaviour was useful, because it provides insights into the visual (where do drivers look), cognitive (when has the drivers cognitively processed the situation enough to select an action and physical processes (how do drivers prepare for action and how do they act) over time. The breakdown of the response into different steps, offers a unique opportunity to understand how these steps interact and which of the steps are affected by the experimental conditions. This understanding is an important step towards identifying the mechanisms that influence driver conflict response in automation.

CRediT authorship contribution statement

Linda Pipkorn: Data curation, Formal analysis, Visualization, Writing - original draft. **Trent W. Victor:** Supervision, Conceptualization, Methodology, Writing - review & editing. **Marco Dozza:** Supervision, Funding acquisition, Writing - review & editing. **Emma Tivesten:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Swedish FFI project ADEST, grant number 2014-06012, within the Drive Me programme and by the European Project L3Pilot, grant number 723051. The authors would like to thank Elin Meltzer, Joel Johansson, Pär Gustavsson and Regina Johansson for assistance with data collection and Giulio Bianchi Piccinini for discussions about, and assistance with, the Predictive Processing framework.

References

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129–135). Pergamon.

Benderius, O. (2014). *Modelling driver steering and neuromuscular behaviour*. Chalmers University of Technology.

Bianchi Piccinini, G., Lehtonen, E., Forcolin, F., Engström, J., Albers, D., Markkula, G., ... Sandin, J. (2019). How do drivers respond to silent automation failures? Driving simulator study and comparison of computational driver braking models. *Human Factors*. 0018720819875347.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381–394.

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G. B., & Victor, T. (2018). Great expectations: A predictive processing account of automobile driving. *Theoretical Issues in Ergonomics Science, 19*(2), 156–194.

Euro NCAP (2018). 2018 automated driving tests: Cut-out scenario. Retrieved from https://www.euroncap.com/en/vehicle-safety/safety-campaigns/2018-automated-driving-tests/.

Gustavsson, P., Victor, T. W., Johansson, J., Tivesten, E., Johansson, R., & Ljung Aust, M. (2018). What were they thinking? Subjective experiences associated with automation expectation mismatch. In *Proceedings of the 6th driver distraction and inattention conference* (pp. 15–17).

Habibovic, A., Andersson, J., Nilsson, M., Lundgren, V. M., & Nilsson, J. (2016). Evaluating interactions with non-existing automated vehicles: three Wizard of Oz approaches. In *2016 IEEE intelligent vehicles symposium (IV)* (pp. 32–37). IEEE.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics, 66*, 18–31.

Larsson, A. F., Kircher, K., & Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 229–237.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). Human-automation interaction. In Designing for people: An introduction to human factors engineering (Chap. 11, pp. 357–388).

Llaneras, R. E., Cannon, B. R., & Green, C. A. (2017). Strategies to assist drivers in remaining attentive while under partially automated driving: Verification of human–machine interface concepts. *Transportation Research Record, 2663*(1), 20–26.

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis & Prevention, 108*, 9–18.

McDonald, A. D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. (2019). Toward computational simulations of behavior during automated driving takeovers: A review of the empirical and modeling literatures. *Human Factors, 61*(4), 642–688.

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., ... McGehee, D. (2019). The "out-of-the-loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work, 21*(1), 87–98.

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In *Human behavior and traffic safety* (pp. 485–524). Boston, MA: Springer.

Mole, C. D., Lappi, O., Giles, O., Markkula, G., Mars, F., & Wilkie, R. M. (2019). Getting back into the loop: The perceptual-motor determinants of successful transitions out of automated driving. *Human Factors*. 0018720819829594.

NHTSA (2015). Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. Retrieved from: https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812115.

NTSB (2017). Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida, May 7, 2016. Retrieved from: https://www.ntsb.gov/investigations/accidentreports/pages/har1702.aspx.

NTSB (2018). Preliminary report highway HWY18FH011. Retrieved from: https://www.ntsb.gov/investigations/accidentreports/pages/hwy18fh011-preliminary.aspx.

NTSB (2019). Preliminary report highway HWY19FH008. Retrieved from: https://www.ntsb.gov/investigations/accidentreports/pages/hwy19fh008-preliminary-report.aspx.

Naujoks, F., Purucker, C., Neukum, A., Wolter, S., & Steiger, R. (2015). Controllability of partially automated driving functions–does it matter whether drivers are allowed to take their hands off the steering wheel?. *Transportation Research Part F: Traffic Psychology and Behaviour, 35*, 185–198.

Naujoks, F., Purucker, C., Wiedemann, K., Neukum, A., Wolter, S., & Steiger, R. (2017). Driving performance at lateral system limits during partially automated driving. *Accident Analysis & Prevention, 108*, 147–162.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253.

Seppelt, B. D., & Victor, T. W. (2016). Potential solutions to human factors challenges in road vehicle automation. *Road vehicle automation* (3, pp. 131–148). Cham: Springer.

Smith, B. W. (2013). Automated vehicles are probably legal in the United States. *Texas A&M Law Review, 1*, 411.

Strand, N., Nilsson, J., Karlsson, I. M., & Nilsson, L. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 218–228.

Tivesten, E., Victor, T. W., Gustavsson, P., Johansson, J., & Ljung Aust, M. (2019). Out-of-the-loop crash prediction: The automation expectation mismatch (AEM) algorithm. *IET Intelligent Transport Systems*.

UNECE (2017). Addendum 78: UN Regulation 79. Retrieved from: https://www.unece.org/fileadmin/DAM/trans/main/wp29/.../R079r4e.pdf.

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. *Human Factors, 60*(8), 1095–1116.

Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009). Identifying black swans in NextGen: Predicting human performance in off-nominal conditions. *Human Factors, 51*(5), 638–651.