Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation

Tyron Louw and Natasha Merat*

Institute for Transport Studies, University of Leeds

ABSTRACT

This driving simulator study, conducted as part of the EC-funded AdaptIVe project, assessed drivers' visual attention distribution during automation and on approach to a critical event, and examined whether such attention changes following repeated exposure to an impending collision. Measures of drivers' horizontal and vertical gaze dispersion during both conventional and automated (SAE Level 2) driving were compared on approach to such critical events. Using a between-participant design, 60 drivers (15 in each group) experienced automation with one of four screen manipulations: 1) *no manipulation*, 2) *manipulation by light fog, 3) manipulation by heavy fog*, and 4) *manipulation by heavy fog with a secondary task*, which were used to induce varying levels of engagement with the driving task. Results showed that, during automation, drivers' horizontal gaze was generally more dispersed than that observed during manual driving. Drivers clearly looked around more when their view of the driving scene was completely blocked by an opaque screen in the *heavy fog* condition. By contrast, horizontal gaze dispersion was (unsurprisingly) more concentrated when drivers performed a visual secondary task, which was overlaid on the opaque screen. However, once the manipulations ceased and an uncertainty alert captured drivers' attention towards an impending incident, a similar gaze pattern was found for all drivers, with no carry-over effects observed after the screen manipulations. Results showed that drivers' understanding of the automated system increased as time progressed, and that scenarios that encourage driver gaze towards the road centre are more likely to increase situation awareness during high levels of automation.

Highlights

- Drivers have more dispersed gaze during automation
- Drivers' visual attention recovers quickly after short periods out-of-the-loop
- Drivers' understanding of an automated driving system increases as time progresses

1. Introduction

The past decade has seen a rapid development of vehicles equipped with Advanced Driver Assistance Systems (ADAS), culminating in multiple vehicle manufacturers releasing first-generation automated driving functionalities such as Lane Keeping Assist (LKA) and Adaptive Cruise Control (Level 2, partial automation; SAE, 2014). These include the Volvo XC90 (Volvo Cars, 2015), Tesla Model S (Tesla Motors, 2015), and Infinity Q50 (Infinity, 2015). While vehicle automation promises a number of social and individual benefits, including increased mobility (Rosenbloom, 2012), safety and efficiency (Anderson et al., 2014), it also shifts the driver's role, from that of an active operator to that of a passive supervisor (Merat, Jamson, Lai, & Carsten, 2012). Some authors have suggested that this supervisory role takes drivers "out-of-the-loop" (OOTL) and impairs their ability to manage critical situations when performance after automation failure/limitations is compared to manual driving (Rudwin-Brown & Parker, 2004; Gold, Damböck, Lorenz, & Bengler, 2013; Strand, Nilsson, Karlsson, & Nilsson, 2014; Merat, Jamson, Lai, Daly, & Carsten, 2014). While the origin of this OOTL concept is based on the effect of automation on performance within other domains (Weiner & Curry, 1980; Bainbridge, 1983; Norman & Orlady, 1989; Endsley & Kiris, 1995; Rasmussen & Rouse, 2013), the term is not yet currently well-defined when addressing the impact of vehicle automation on driving performance. Yet, from a human factors and road safety perspective, it is important to investigate the nature and consequences of this OOTL state and understand, for example, how it influences drivers' distribution of attention during high levels of automation, or how it affects their ability to resume control from automation in an appropriate and timely manner, should a system limit be reached. This paper, therefore, describes a driving simulator study that attempted to simulate the OOTL concept in vehicle automation and reports on the distribution of drivers' visual attention during SAE level 2 automation as a means of assessing this methodology.

According to Kienle et al. (2009), a driver is considered OOTL when they are "not immediately aware of the vehicle and the road traffic situation because they are not actively monitoring, making decisions or providing input to the driving task". Norman (1990) attributes causality not to automation per se but rather to a lack of continual feedback. The concept seems, therefore, to include two elements; one, which relates to the awareness of elements in the environment, and another, which relates to the awareness of elements regarding vehicle status and its automated system(s).

Seeking to expand on the mechanisms underlying the OOTL problem, Louw, Kountouriotis, Carsten, & Merat (2015) presented a schematic representation of this concept, which proposes that, as a result of vehicle automation, drivers are removed from a physical control loop, because they are no longer physically interacting with the vehicle's mechanisms such as the steering wheel and pedals (see also Stanton & Young, 1998). Drivers can also be removed from a 'cognitive control loop' and lose situation awareness, either because they are looking away from the driving scene during automation and interacting with a distracting task, or due to boredom/mind-wandering (Lerner et al., 2015). Clearly, both loops are important for contributing to safe driving performance, since, for instance, physical neuromuscular control gives drivers feedback of steering torque and helps contribute to corrections of heading errors (Pick & Cole, 2006), whilst good situation awareness contributes to effective attentional control and decision-making and improves hazard perception, for instance, in response to critical events (Endsley, 2006; Horswill & McKenna, 2004). Accordingly, Louw et al. (2015) hypothesise that reductions in either or both aspects of control, brought about by automation, can contribute to less effective return-to-manual performance, but that not being in physical control can also act to impair situation awareness, which consequently can reduce driving performance.

To further investigate this concept, the current study sought to induce a range of OOTL states by removing driving-relevant information during automation and explored whether these affected drivers' ability to regain situation awareness in response to a potentially critical event. Based on the Kienle et al. (2009) definition, being in the loop involves three distinct elements: drivers must (i) be aware of the vehicle (ii) be aware of the road traffic situation and (iii) make decisions or provide input to the driving task (when resuming control). We, therefore, designed a study where we examined how drivers' ability to respond to potentially critical situations which followed a system-initiated automation disengagement, was affected by the systematic removal of the three elements mentioned above, thereby inducing an artificial OOTL state. This was achieved by developing a screen manipulation technique, introduced in Louw et al. (2015) and

Louw et al. (2016), which uses a fog-like display to vary the degree of visual information available to drivers during automation, both in terms of the dashboard displays in the vehicle and also the road environment itself (see Figure 2, and Methods section for a more detailed outline). This approach broadly resembles a visual occlusion technique, first used by Senders et al. (1967) to model driver behaviour based on information theory, and then others to quantify the visual demand of in-vehicle information systems (Foley, 2008).

Extended durations of automated driving have been shown to take drivers further OOTL (Körber, Cingel, Zimmermann, & Bengler, 2015). However, here, we were simply interested in assessing whether removing driving-relevant information, with short periods of such screen manipulations, would take drivers OOTL, and what the effects of such manipulations would be on drivers' visual attention. Of course, one simple method for taking drivers OOTL (both physical and cognitive) is to allow interaction with a secondary task during automation. However, our rationale for using screen manipulations was to reduce the complications associated with the physical demand of engaging in a secondary task (Zeeb, Buchner, & Schrauf, 2015), which can take drivers' head, hands and eyes away from the driving scene (Carsten, Lai, Barnard, Jamson, & Merat, 2012; Louw, Merat, & Jamson, 2015) and adds considerable individual variability during the return to manual control.

Traditionally, analysis of drivers' performance in the transition period from automation to manual control has relied on the use of vehicle-based metrics and reaction time measures, following a mandatory resumption of control from a failing or limited automation system (Gold, Damböck, Lorenz, & Bengler, 2013; Louw et al., 2015; Merat & Jamson, 2008). However, while it is relevant to establish the minimum time required for drivers to resume control of the vehicle after automation disengagement (termed a take-over-response or TOR; see Beller, Heesen, & Vollrath, 2014; and Helldin, Falkman, Riveiro, & Davidsson, 2013), we argue that such instructions to resume control may simply be in response to alarms and experimenter commands, and not a reflection of drivers' recognition of, and ability to manage, an emerging critical situation. This argument is supported by Gold and colleagues' finding that while a relatively rapid resumption of control from automation is possible, where the first braking input can be as fast as 2.06s, and steering input is around 2.27s, it is at the cost of safe vehicle control (Gold, Damböck, Lorenz, & Bengler, 2013). Therefore, our aim was to investigate drivers' assessment of the environment following a period of screen manipulation using an uncertainty alert, which declared the automation might not be able to handle the unfolding situation, and investigated how each screen manipulation condition affected drivers' ability to evaluate the criticality of events and decide whether resumption of control was necessary. We also assessedwhether repeated exposure to such events influenced drivers' visual attention.

To assess drivers' attention to the driving scene and vehicle controls during, before and after each screen manipulation, we considered their visual attention to different areas of interest, using eye gaze dispersion. Psychophysiological research using eye gaze data has been a popular method for measuring drivers' attention allocation (Posner, 1980), situation awareness (Gugerty, 2011; Gartenberg, Breslow, McCurry, & Trafton, 2013) and hazard perception (Endsley & Jones, 2004; Horswill & McKenna, 2004). However, while gaze concentration has been used successfully in manual driving to distinguish between the effects of visual and cognitive load (Engström, Johansson, & Ostlund, 2005), it has been scarcely applied in automated driving (see for example Damböck, Weißgerber, Kienle, & Bengler, 2013, who report greater horizontal gaze dispersion for highly automated driving as compared to manual driving). A review of the literature by De Winter, Happee, Martens, & Stanton (2014), found that drivers in highly automated driving gaze on the road less often than when in manual control, which therefore could result in lower workload, but also poor situation awareness. However, most of the studies reviewed by De Winter et al. (2014) have used fixation-based Percentage Road Centre (PRC) measures (e.g. Carsten et al., 2012), rather than raw gaze data. According to Wang, Reimer, Dobres, & Mehler (2014), fixation-based PRC is less sensitive to demand-induced changes in visual behaviour than measures of gaze-based PRC and gaze dispersion. Therefore, in this study, we chose to explore the use of horizontal and vertical gaze dispersion as a means of evaluating drivers' OOTL state during automated driving, as well as during the resumption of control from automation. The screen manipulation technique was used to induce varying levels of the OOTL state, by systematically removing information from drivers during automation. The study then considered the following questions:

i. What gaze pattern do drivers exhibit during each of the different screen manipulation conditions?
ii. When resumption of manual control is required, is drivers' visual attention to the scene and vehicle controls affected differently by the different screen manipulations?
iii. Can we infer drivers are taken out of the loop by the screen manipulations, and does this depend on the particular manipulations applied?
iv. Does drivers' visual attention change after repeated exposure to the same events?

2. Methods

2.1. *Participants*

Following approval from the University of Leeds Research Ethics Committee (Reference Number: LTTRAN-054), four groups of 15 drivers were recruited via the driving simulator database and were paid £20 for taking part in the experiment. The average age of the participants was 36.16 ± 12.38 years, and out of 60 participants, 32 were male. Average mean annual mileage was 8290.46 ± 6723.08 miles. Participants had normal or corrected-to-normal vision, were required to have had a driving licence for at least one year (M = 16.22, SD = 12.92) and drive at least twice a week. Data from one participant was excluded from the analysis due to abnormal values from their eye-tracking data (±3 SD from the mean).

2.2. *Design and Procedure*

2.2.1. *Materials*

The experiment was conducted in the University of Leeds Driving Simulator, which consists of a Jaguar S-type cab with all driver controls operational. The vehicle is housed in a 4m spherical projection dome and has a 300° field-of-view projection system. A v4.5 Seeing Machines faceLAB eye-tracker was used to record eye movements at 60Hz.

2.2.2. *Design*

A repeated measures mixed design was used for this study, with a between-participant factor of Screen Manipulation (no fog, light fog, heavy fog, heavy fog + task) and within-participant factors of Drive Type (manual, automated) and Event Number (1-6).

The experimental session consisted of two drives for each group (manual, automated) which lasted about 20 minutes each, and participants experienced a short break between drives, to alleviate the symptoms of fatigue. Participants drove the same road in both drives, but the screen manipulation was only used during the automated drives. For each Screen Manipulation group, the order of drives was counterbalanced across participants. As shown in Figure 1, within each automation and manual drive, there were six discrete car-following events, each lasting approximately 150 s. Our main aim here was to study drivers' response to critical events after they were taken OOTL with a screen manipulation. However, to assess situation awareness after the uncertainty events (see below), and to reduce priming, each drive contained only two critical events (events 2 and 6), interspersed with four non-critical events (events 1, 3, 4 and 5). During the non-critical events, the lead vehicle would either speed up or change lane, while during the critical events the lead vehicle decelerated at a rate of $5 m/s^2$, resulting in an impending collision scenario. The time-to-collision (TTC) at the start of this deceleration was 3 s.

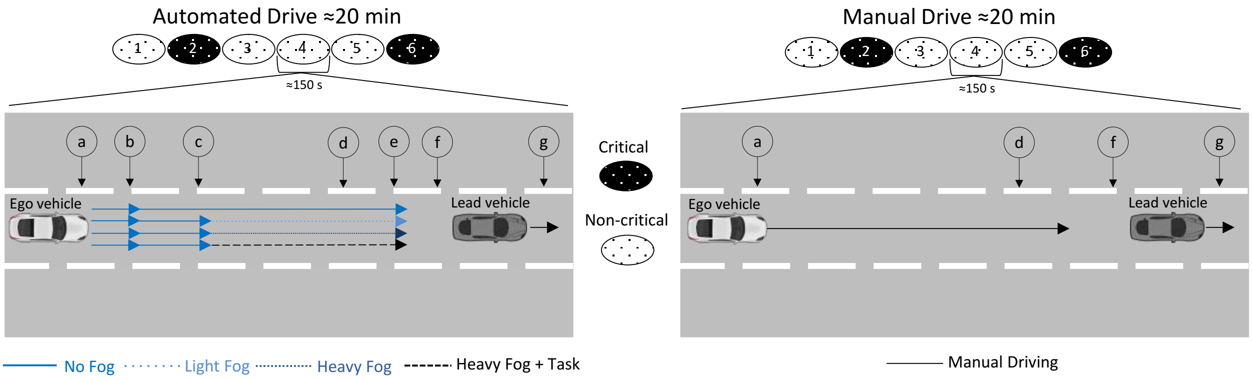

Figure 1. Schematic representation of each discrete event in the automated (left) and manual (right) drives. (a) to (g) represent various phases of the drive, as follows: (a) Event start, (b) Automation on, (c) Screen Manipulation on, (d) Drone moves into lane, (e) Screen Manipulations off + uncertainty alert, (f) Drone action, (g) Event end.

As outlined in the Introduction, to induce varying levels of the OOTL state during the automated drives, we employed four screen manipulation techniques (Figure 2). In the no fog condition, there was no manipulation of the road scene, and drivers could observed all aspects of the road and traffic environment. In the light fog condition, a translucent grey filter superimposed the road scene. The aim of this manipulation was to simulate a process whereby drivers were able to distinguish only basic elements of the road environment and the movement of vehicles in the immediate vicinity. In the heavy fog condition, an opaque grey filter overlaid the road scene. This manipulation sought to effectively blocked all visual information from the road environment such that drivers were unaware of the traffic conditions. During the heavy fog + task condition the road was blocked with the same opaque grey filter used in the heavy fog condition but overlaid with a series of visually presented secondary tasks. Here, participants were required to complete a number of multiple-choice questions involving visuospatial shape-matching, general knowledge questions, and moderately challenging mathematical questions, which were sourced from various web-based IQ tests and were presented in a random order. All responses to this task were verbal. The aim of this manipulation was to assess how engagement in a secondary task affected performance, but since we were keen not to remove drivers' eyes and head away from the screen (keeping physical position as similar as possible to the other experiments) the secondary task was displayed on the driving scene, akin to a Head-Up Display. Participants were told that they would not be penalised for incorrect answers, but that their response would be recorded. We hypothesised that less visual information about the scene would take drivers further OOTL and that therefore drivers were most OOTL during the heavy fog condition, followed by light fog and no fog.

2.2.1. *Procedure*

Upon arrival, participants were briefed on the description of the study and were asked to sign a consent form, with an opportunity to ask any questions, if required. They were then given the chance to practice manual driving and Highly Automated Driving (HAD) within a free-flowing three-lane motorway. During the practice session, participants were talked through the various aspects of the vehicle HMI (Figure 3), were shown how to engage and disengage the automation and were shown the screen manipulation they would encounter during the experimental automated drive. The road contained ambient traffic, but participants did not experience the critical events during the practice drives.

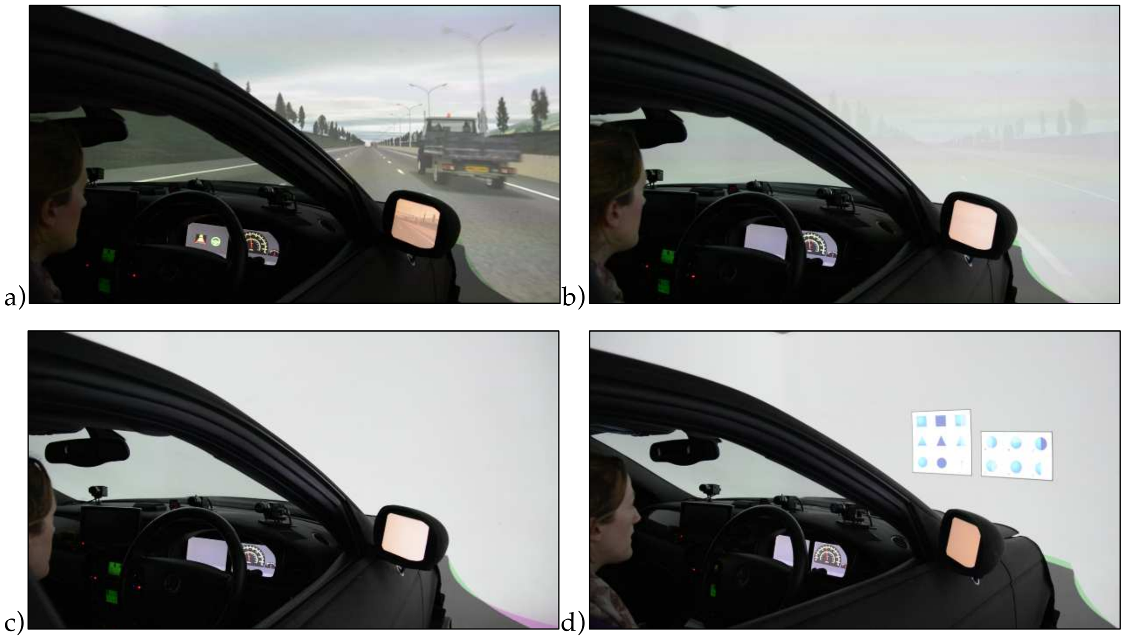

Figure 2. Example of a drivers' view in the a) no fog, b) light fog, c) heavy fog, and d) heavy fog + task conditions

Regarding automation uncertainty, participants were told that if the automation uncertainty HMI appeared (see below for how this was portrayed), they should monitor the driving environment and determine for themselves whether or not to intervene. Participants were instructed to drive in the middle lane of the three-lane motorway for the duration of the drive (automation was only possible in this lane) but were permitted to change lane in critical situations, and were told to move back into the middle lane as soon as possible. Drivers were otherwise asked to obey the standard rules of the road and to ensure safe operation of the vehicle.

To engage the highly automated driving system, participants pressed a button on the steering wheel. To disengage automation, participants would either press the same button, turn the steering wheel more than 2° or press the brake pedal. During the automated drive, participants were asked to move to the centre of the middle lane as soon as convenient and then activate automated driving as soon as it was available. If drivers did not engage automation, the system engaged automatically after 5 s. The activation of automation constituted the start of an event. After 30 s of automated driving, one of four 90 s screen manipulations began. It is important to note that the vehicle dynamics, as well as all auditory cues, remained active during the screen manipulations. After each screen manipulation, the presence of a lead vehicle triggered an uncertainty scenario (for both critical and non-critical events). At this point, the screen manipulation concluded, the driving scene was again visible, and the automation status changed from "Engaged" to "Uncertain". Drivers were notified of this change by a short duration auditory tone (1000Hz, lasting 0.2 s), and the automation status symbol, which was now visible, changed from green to flashing yellow. The driver was expected to monitor the driving situation and intervene, if necessary. After 3 s, the lead vehicle completed one of three manoeuvres: In the non-critical event (1, 3, 4, 5) the lead vehicle either moved out of lane 2 or sped up, while in the critical events (2, 6) the lead vehicle braked sharply with a maximum deceleration of $5.0\ m/s^2$.

2.2.2. *Human-Machine Interface (HMI)*

The status of the vehicle's automated system was indicated by the colour of a steering wheel symbol that was located on the left panel of the central display unit (Figure 3). During the automated drives, the steering wheel symbol was solid green when automation was engaged, flashing yellow when it was uncertain and solid grey when it was unavailable. Any change to the automation state, whether driver- or system-initiated, was accompanied by the same non-intrusive auditory tone described above.

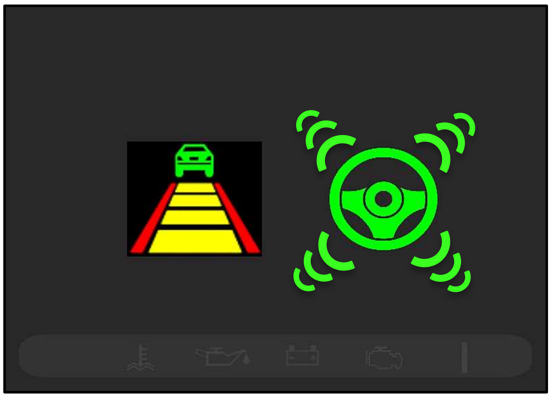

Figure 3. An example of the in-vehicle HMI with the Forward Collision Warning symbol on the left and the Automation Status Symbol on the right (flashing green in this example).

In addition to the automation status, a Forward Collision Warning (FCW) symbol was included in the left panel of the central display unit. Active only when automation was engaged, this system provided a visual approximation of the headway of the lead vehicle in seconds. In the automated drives, a continuous alarm alerted drivers of an imminent collision whenever TTC with the lead vehicle was below a 2 s threshold. However, this only occurred during the critical events. To further deprive drivers of system information during automation, the automation status (steering wheel) and the FCW were also hidden during the screen manipulation conditions. However, participants were able to reveal the HMI at any point by pulling the left indicator stick towards them. This action illuminated the HMI for 2-seconds. Participants were able to move this stick as often as they wished.

2.3. *Statistical analyses*

All data were analysed with IBM SPSS v21 (IBM Corp., 2012). Shapiro Wilk's test showed that not all estimates were normally distributed. As the data were moderately positively skewed square root transformations were used for analyses (Tabachnick and Fidell, 2007). ANOVA results reported below are based on the transformed responses, while the graphs represent estimates in the original units, to facilitate interpretation (Neter et al., 1990). An α-value of .05 was used as the criterion for statistical significance and partial eta-squared was computed as effect size statistics. Degrees of freedom were Greenhouse-Geiser corrected when Mauchly's test showed a violation of sphericity. Unless otherwise stated, variances of the data were homogenous, as assessed by Levene's test of equality of error variances (Field, 2009). Similarly, covariances of the data were homogenous, as assessed by Box's test of equality of covariance matrices, unless otherwise stated. LSD pairwise comparisons (α = .05) were used to determine the difference between levels of Screen Manipulation and Event Number.

As highlighted in the Introduction, we used drivers' gaze dispersion to establish how visual attention was distributed before and during each of the different screen manipulations. In addition, to understand how each manipulation affected this dispersion on approach to the six events, we considered how gaze dispersion varied just after the screen manipulations. To compare across the groups, we, therefore, divided each event into five Time Windows (TW), as in Figure 4. The rationale for these divisions and analyses are summarised in Table 1. Full statistical results are included in Table 2.

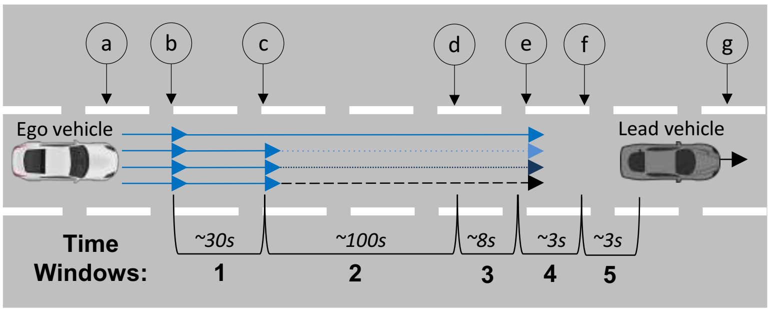

Figure 4. Schematic representation of the Time Windows used for the analyses. (a) to (g) represent various phases of the drive, as follows: (a) Event start, (b) Automation on, (c) Screen Manipulation on, (d) Drone moves into lane, (e) Screen Manipulations off + uncertainty alert, (f) Drone action, (g) Event end.

3. **Results and Discussion**

3.1. *Gaze patterns during uninterrupted driving (Time Window 1)*

Time Window 1 (TW1) was the only period in the automated drive where all drivers were able to see the road environment, which therefore allowed a comparison of performance with manual driving, and provided a reference point for the four screen manipulation conditions. For SD of Gaze Yaw, a three-way ANOVA showed a significant effect of Drive Type, where horizontal scanning was higher during automated driving compared to manual driving (M = 8.35°, SEM = .39° vs. M = 6.92°, SEM = .29°, respectively; Table 2). Our results are in line with findings from Damböck et al. (2013) and multiple other studies, which have used gaze PRC (De Winter et al., 2014) and find higher horizontal scanning by drivers during automation. This pattern of increased horizontal scanning can be seen in Figure 5, which shows an example of density contour plots of gaze dispersion for the automated and manual drives during TW1, on approach to a non-critical event (Event 5). Analyses of variance did not find a significant effect of Screen Manipulation or Event Number for SD of Gaze Yaw, or any significant interactions.

Table 1 - Time windows used for statistical analyses

Time Window	Start	End	Rationale	Comparisons
1	Automation On (b)	Screen Manipulation On (c)	Assess visual attention during uninterrupted driving	2 X 6 X 4 ANOVA: Drive Type (*automated, manual*) and Event Number (*1-6*) as within-participant factors and Screen Manipulation (*No Fog, Light Fog, Heavy Fog, Heavy Fog + Task*) as a between-participant factor
2	Screen Manipulation On (c)	Drone Moves Into Lane (d)	Assess the effect of the screen manipulations on visual attention	6 X 4 ANOVA: Event Number (*1-6*) as within-participant factors and Screen Manipulation (*No Fog, Light Fog, Heavy Fog, Heavy Fog + Task*) as a between-participant factor
3	Drone Moves Into Lane (d)	Screen Manipulation Off (e)	Not used for analysis	-
4	Screen Manipulation Off (e)	Lead Vehicle Action (f)	Assess the carry-over effect of the screen manipulations on visual attention	2 X 6 X 4 ANOVA: Drive Type (*automated, manual*) and Event Number (*1-6*) as within-participant factors and Screen Manipulation (*No Fog, Light Fog, Heavy Fog, Heavy Fog + Task*) as a between-participant factor
				6 X 4 ANOVA: Event Number (*1-6*) as within-participant factors and Screen Manipulation (*No Fog, Light Fog, Heavy Fog, Heavy Fog + Task*) as a between-participant factor
5	Lead Vehicle Action (f)	Lead Vehicle Action (f) + 3 s	Assess the effect of a lead vehicle braking on visual attention	2 X 4 ANOVA: Critical Event (*2, 6*) as within-participant factors and Screen Manipulation (*No Fog, Light Fog, Heavy Fog, Heavy Fog + Task*) as a between-participant factor

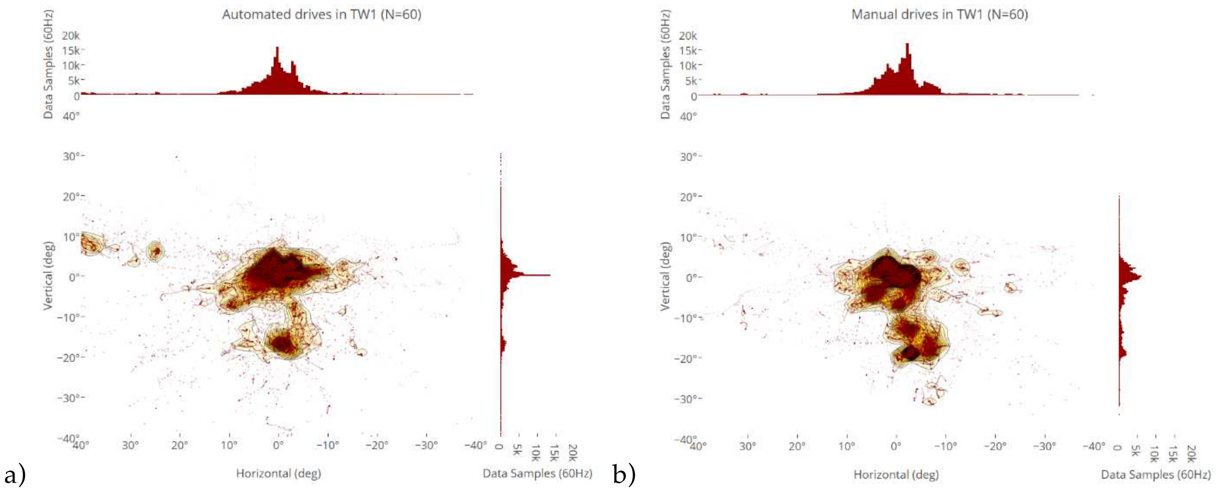

Figure 5. Example density contour plots of gaze dispersion in Time Window 1 (from "Automation On" to "Screen Manipulation On") for Event 5 for the a) automated and b) manual drives. The primary plots illustrate a 40° vertical and horizontal field of view, where darker areas represent more concentrated gaze areas, while the histograms depict two-dimensional views of horizontal gaze concentration and the histograms to the right depict two-dimensional views of vertical gaze concentrations.

For SD of Gaze Pitch, a three-way ANOVA revealed no effect of Screen Manipulation or Drive Type (Table 2). There was a significant effect of Event Number for SD of Gaze Pitch, with Figure 6 showing that there was a gradual decrease across the experiment for both drives. However, the significant interaction of Drive Type and Event Number, also shown in Figure 6, suggests that the effect of Event Number is mainly due to the automated drives, as post-hoc tests showed that vertical gaze was significantly more dispersed in Event 1 compared to Events 2-5 ($p<.001$). This higher SD of Gaze Pitch for the first Event in automation was likely due to a familiarisation period, as drivers tried to assess their environment, looking ahead at the driving scene and back at the vehicle dashboard, which is then significantly reduced after the first Event. Also, whereas in the automated drive SD of Gaze Pitch continues to fluctuate over the course of the six events, in the manual drive the decrease is relatively consistent, which is likely because, across all time windows, the manual drive was far less interrupted. Therefore, drivers seem to be looking between the lead vehicle and dashboard less during automation, perhaps attempting to assess the environment they do not control. In manual driving, drivers divided their attention between the lead vehicle and dashboard more, which could be because they were asked to maintain a speed of 70mph.

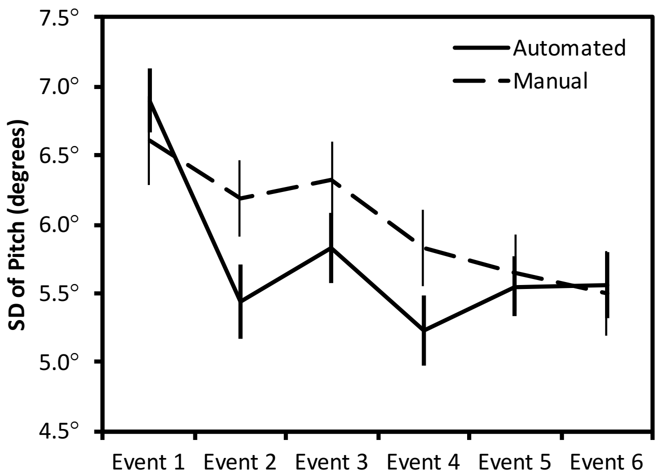

Figure 6. Mean SD of Gaze Pitch in Time Window 1 (from "Automation On" to "Screen Manipulation On") for each Event Number for the automated and manual drives.

The slight increase in SD of Gaze Pitch in Event 3 for both the automated and manual drives is likely due to drivers' propensity to engage further in the driving task and glance more regularly between the road and vehicle console after experiencing the first critical event (Event 2).

3.2. *Gaze patterns during the screen manipulations (Time Window 2)*

Time Window 2 (TW2) represents the period where the various screen manipulations were applied during the automated drives. Here, we expected changes to visual attention distribution as a result of these manipulations, and as a direct consequence of the degree of visual information available to drivers. To provide an overview of gaze distribution for the four Screen Manipulation conditions, Figure 7 displays combined density contour plots for all participants during TW2 for each Manipulation.

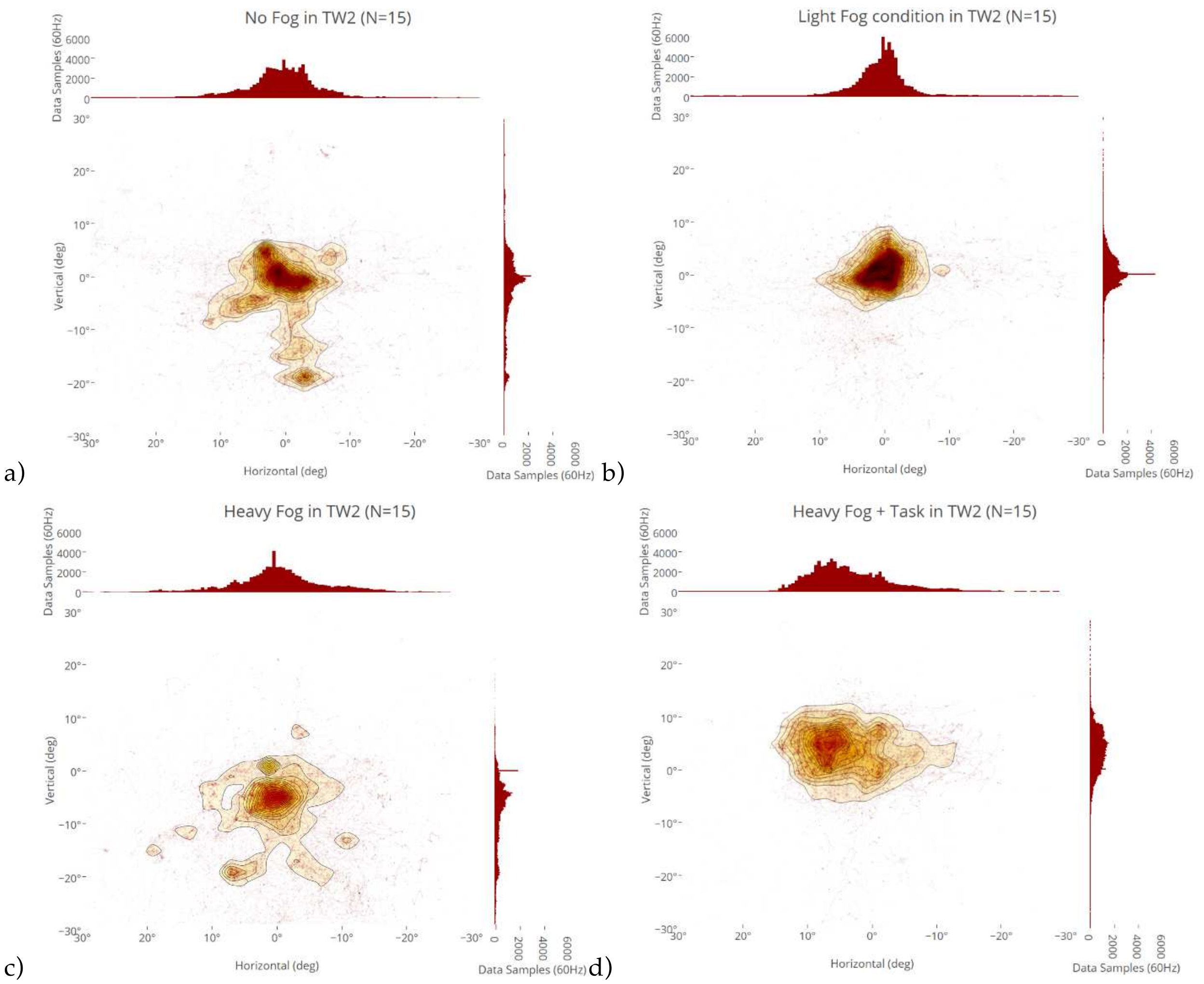

Figure 7. Density contour plots of gaze for the a) *no fog*, b) *light fog*, c) *heavy fog* and d) *heavy fog + task* conditions in Time Window 2 (from "Screen Manipulation On" to "Drone Moves into Lane") for all Event Numbers in the automated drives.

A two-way ANOVA was conducted on the SD of Gaze Yaw to assess the effect of each screen manipulation during automation (Table 2). Results showed a significant main effect of Screen Manipulation, as illustrated in Figure 8. Post-hoc analyses revealed this was due to the different horizontal scanning pattern of drivers during the no fog and heavy fog + task conditions ($p < .05$). This reduction of horizontal gaze because of the secondary task engagement is not surprising and cannot in itself be used as a direct indication of how much drivers were OOTL. Figure 8 also shows similar horizontal scanning when drivers were able to see the driving scene in the no fog condition and when the scene was fully occluded during the heavy fog condition. There was no effect of Event Number ($p = .664$) for SD of Gaze Yaw and no interaction between Event Number and Screen Manipulation, suggesting that the screen manipulations had a consistent effect on horizontal scanning throughout the six Events for all screen manipulations ($p = .92$).

A two-way ANOVA for SD of Gaze Pitch also showed a main effect of Screen Manipulation, with the greatest vertical gaze dispersion seen for drivers in the heavy fog condition. Post-hoc analyses found significant differences between this condition and all other screen manipulation conditions (Figure 8), suggesting that when drivers were taken OOTL by not being able to see the road, their primary vertical gaze activity focused on looking between the road ahead and the vehicle dashboard, presumably awaiting the end of the screen manipulation. There was also a main effect of Event Number for SD of Gaze Pitch. As can be seen in Figure 9, pairwise comparisons revealed that this effect was due to an increased concentration of vertical scanning after Event 1, which is significantly higher than all but the last Event (p<.05). As with TW1, the higher SD of Gaze Pitch for Event 1 is likely due to a familiarisation period by drivers, at the start of the drive. There was no interaction between Event Number and Screen Manipulation for SD of Gaze Pitch.

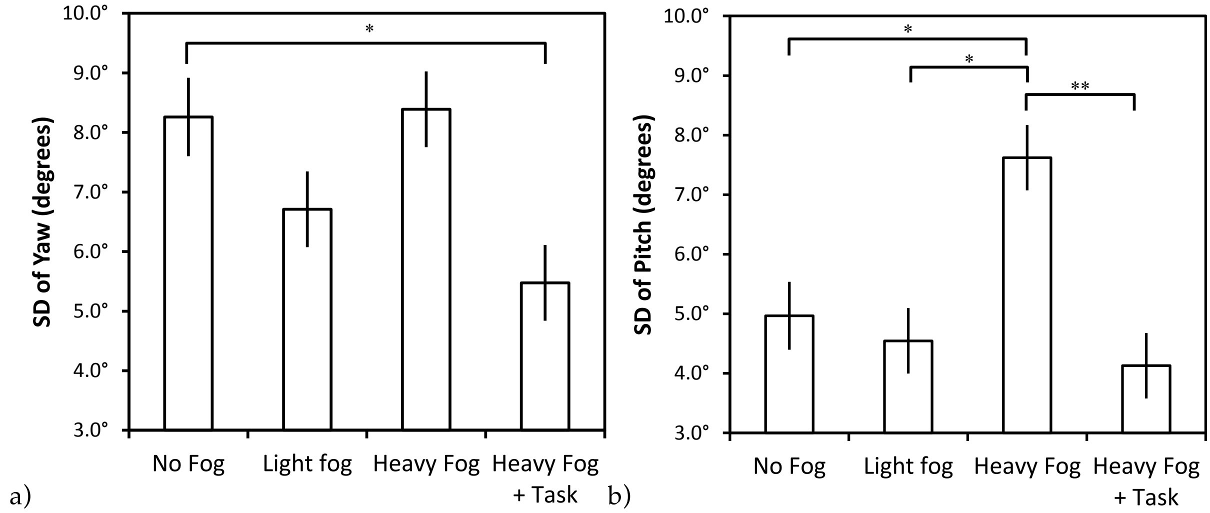

Figure 8. Mean SD of a) Yaw and b) Pitch during Time Window 2 (from "Screen Manipulation On" to "Drone Moves into Lane") for each of the four automated drives (*p<.05, **p<.001).

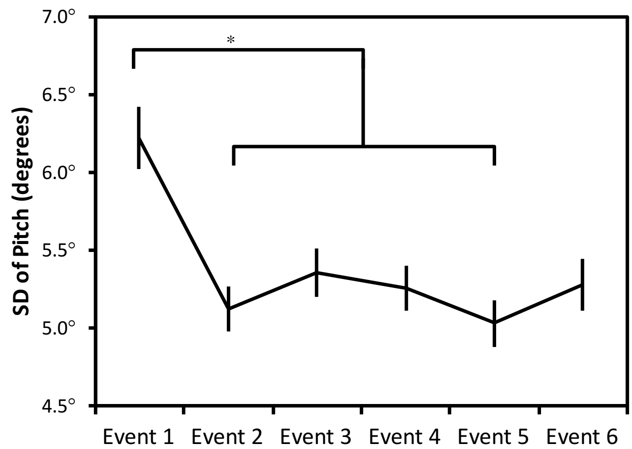

Figure 9. SD of Gaze Pitch across all events in the automated drive for Time Window 2 (from "Screen Manipulation On" to "Drone Moves into Lane"). Asterisks indicate that, for SD of Gaze Pitch in the automated drive, Event 1 is significantly different to Events 2-5 (*p<.05).

3.3. *Gaze patterns pre-screen manipulations (Time Window 3)*

Time Window 3 (TW3) constituted an 8 s period where the manipulations in TW2 continued in the automated drives, but where surrounding vehicles began to move into place to trigger an uncertainty event. TW3 was excluded from the analyses since drivers' eye movement were likely to be affected by the movement of surrounding vehicles when the road scene was visible.

3.4. *Gaze patterns post-screen manipulations (Time Window 4)*

Time Window 4 (TW4) constituted a 3 s period from the end of the screen manipulations (which coincided with the start of the uncertainty alert in automation) up to the moment before the lead vehicle either braked, changed lane, or sped up. This period allowed for the assessment of drivers' gaze patterns during a Situation Awareness Recovery period (SAR; Gartenburg et al., 2013), defined as the process of restoring SA after SA has been reduced. A 3-way ANOVA revealed no differences between the automated and manual drives for either SD of Gaze Yaw ($p = .130$) or SD of Gaze Pitch ($p = .160$), suggesting that regardless of the screen manipulation, drivers recovered quite quickly, at least as indicated by their visual attention to the road ahead.

To investigate any carry-over effects of the four screen manipulations on gaze patterns, a 2-way ANOVA was conducted on SD of Gaze Yaw and Pitch for the automated drives only. Results showed that horizontal and vertical gaze dispersion was the same for all Screen Manipulations, in the three seconds immediately after screen manipulation was removed. These results suggest that when drivers' attention was captured by the uncertainty alert, their visual attention to the road ahead was not affected by the previous screen manipulation. Therefore, regardless of the degree of visual information available to drivers during each screen manipulation the same pattern of horizontal and vertical gaze scanning was observed in preparation for response to the lead vehicle.

Analyses of variance also showed that while there was no effect of Event Number for SD of Gaze Yaw in the automated drives ($p = .450$), there was a significant effect of Event Number for SD of Gaze Pitch ($p < .05$). Patterns were similar to that seen during Time Windows 1 and 2 and was likely due to drivers' familiarisation with the driving scenarios, after Event 1.

3.5. *Gaze patterns post-Brake light (Time Window 5)*

Time Window 5 (TW5) constituted a 3 s period after first onset of the lead vehicle's brake light. Only gaze patterns for the two Critical Events were considered for this analysis, as these events required direct intervention by drivers, which would otherwise result in a collision.

A 3-way ANOVA showed that, there was no effect of Drive Type ($p = .225$) or Screen Manipulation ($p = .067$) on SD of Gaze Yaw (Table 2). There was, however, a significant difference in horizontal scanning between the two critical events, with lower SD of Gaze Yaw in the first ($M = 5.26°$, $SEM = .403°$) compared to the second ($M = 6.30°$, $SEM = .51°$) critical event. The ANOVA also revealed a significant interaction between Drive Type and Event Number for SD of Gaze Yaw. Although horizontal scanning was relatively high for both events in the manual conditions, in the automated drive it increased from 4.03° ($SEM = .47°$) in Critical Event 1 to 6.83° ($SEM = .51°$) in Critical Event 2 (Figure 10). This suggests a learning effect in the automated drives, where drivers understood the significance of a potential collision in the first critical event and scanned the environment and particularly the adjacent lane more extensively on approach to the second critical event to prepare for a suitable response, such as changing lane. Clearly, the same degree of horizontal scanning occurred for both Events in manual driving, when drivers were in control of the vehicle and responsible for changing lane.

For SD of Gaze Pitch, there was no effect of Drive Type ($p = .064$) or Screen Manipulation ($p = .095$), suggesting that drivers' vertical gaze distributions just before response to the braking lead vehicle were the same in the manual and automated drives and across the four conditions. There was also no effect of Event Number for SD of Gaze Pitch, suggesting that the screen manipulations did not have a carry-over effect on vertical gaze patterns, once the lead vehicle braked in the critical events.

Table 2 - Results for ANOVAs conducted for each Time Window for SD of Gaze Yaw and SD of Gaze Pitch.

Effect		Drive Type		Event Number		Screen Manipulation		Drive Type*Event Number		Drive Type*Screen Manipulation		Event Number*Screen Manipulation	
		F(df1,df2)	η_p^2	F(df1,df2)	η_p^2	F(df1,df2)	η_p^2	F(df1,df2)	η_p^2	F(df1,df2)	η_p^2	F(df1,df2)	η_p^2
TW1 - 3-way ANOVA	SD of Gaze Yaw	19.9(1,55)**	.266	n.s.		n.s.		n.s.		n.s.		2.0(15,275)*	.099
	SD of Gaze Pitch	n.s.		20.2(4.12,234.81)**	.265	n.s.		2.7(5,280)*	.047	n.s.		n.s.	
TW2 - 2-way ANOVA	SD of Gaze Yaw	N/A		n.s.		4.6(3,56)*	.189	N/A		N/A		n.s.	
	SD of Gaze Pitch	N/A		6.6(3.73,208.79)**	.105	7.6(3,56)**	.298	N/A		N/A		n.s.	
TW4 - 3-way ANOVA	SD of Gaze Yaw	n.s.		2.3(5,280)*	.039	n.s.		n.s.		n.s.		n.s.	
	SD of Gaze Pitch	n.s.		5.7(5,280)**	.093	n.s.		2.9(5,280)*	.049	n.s.		n.s.	
TW4 - 2-way ANOVA	SD of Gaze Yaw	N/A		n.s.		n.s.		N/A		N/A		n.s.	
	SD of Gaze Pitch	N/A		9.4(5,280)**[1]	.143	n.s.		N/A		N/A		n.s.	
TW5 - 3-way ANOVA	SD of Gaze Yaw	n.s.		5.1(1,55)*[1]	.085	n.s.		5.9(1,55)*[1]	.097	n.s.		n.s.	
	SD of Gaze Pitch	n.s.		n.s.		n.s.		n.s.		n.s.		n.s.	

* $p < .05$ ** $p < .01$ [1] Critical Events only (2, 6)

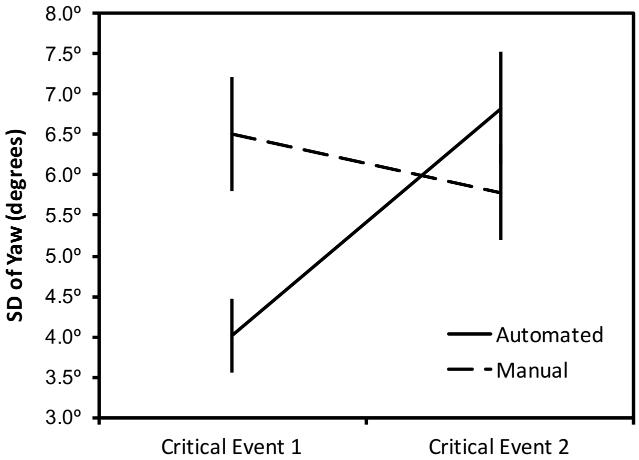

Figure 10. Mean SD of Gaze Yaw for Critical Event 1 and Critical Event 2 in the Automated and Manual drives, for Time Window 5 (from "Drone Action" to "Event End").

4. Conclusions

Previous studies have suggested that drivers' inability to respond effectively to critical scenarios following limitations or failures of highly automated driving (SAE Level 2) is because they are 'out-of-the-loop' (OOTL). In these studies, when performance in automated driving is compared to manual driving, reaction time to critical incidents is slower, sometimes leading to crashes, and drivers are generally less aware of their surroundings, presumably taking some time to reorient their attention to the driving scene after automation is disengaged. However, there is currently no consensus as to what constitutes an OOTL driver, how this state is measured, and what information drivers use to remain engaged with the driving task. To address these issues, we manipulated the simulated driving scene in a series of conditions, by removing driving-relevant visual information for short periods, and investigating driver behaviour and gaze patterns before, during and after such manipulations. Drivers' visual attention to unfolding critical and non-critical events after such manipulations were also studied and findings were compared to driving with manual control.

Results showed that, during automation, drivers' vertical gaze was most dispersed when the road scene and dashboard were completely occluded during automation (heavy fog condition). Here, drivers systematically moved their gaze between the road ahead and the vehicle dashboard, presumably in preparation for the resumption of control. Horizontal gaze dispersion was also highest in this drive. In contrast, and against our expectations, when the road scene was partially occluded during the light fog condition, drivers' gaze was almost entirely on the road centre, and they seemed to ignore the vehicle HMI. Therefore, by withholding only some information, drivers were seen to remain more engaged in the driving task, compared to if all information was removed. Focus of gaze towards the road scene and infrequent gaze towards the HMI was also high when a secondary task was present on the driving scene during automation.

These gaze dispersion patterns provide some understanding of what information drivers use to keep themselves engaged in the driving task during automation, and what information they use to keep in the loop when they have access to only some driving-relevant information. In other words, when they can see the scene for themselves, drivers distribute their gaze towards a larger area of the driving scene and surrounding environment, presumably, because they can easily see any unfolding events and believe they can rapidly resume control from automation, if required. Perhaps they can afford to trust the automation more in this condition as resumption of control is easier, in the event of a failure. When this information is completely removed, attention is also spread horizontally across the driving scene but also heavily between the vehicle HMI and the road. This increase in vertical gaze, especially, is likely to allow drivers access to maximum information from all relevant sources, for resumption of control. Taken together, these results may suggest that, if relevant vehicle information is presented in the dashboard area, then vertical dispersion of gaze is likely to be higher when drivers are OOTL.

When the road scene was only partially visible, drivers clearly believed visual attention was best placed towards the area of the likely incident, the road centre, and that less valuable information was available from the HMI. Here, drivers probably trusted the automated system least and wished to rely on their own skills for resumption of control. Finally, when required to engage in a secondary task, drivers prioritised this task, perhaps to the detriment of driving, also taking their attention away from the HMI – when the road scene was not visible.

Perhaps fortunately, this study revealed that regardless of these screen manipulations, when an uncertainty alert captures drivers' attention towards an impending incident, a similar gaze pattern is seen for all drivers, with no carry-over effects observed after the screen manipulations, and similar gaze patterns also seen for this period of manual driving. Therefore, while these short periods of screen manipulation may well disperse drivers' visual attention away from the road centre, they do not have a long lasting effect on visual attention to the point of danger, when response is required before a potentially critical event.

While the focus of this paper was to induce varying degrees of being OOTL and assess their concomitant effect on drivers' gaze dispersion, the effect of such gaze patterns on ensuing performance is perhaps worth considering. Although not reported here, an analysis of driving performance suggests that those screen manipulations which caused the most gaze dispersion (no fog and heavy fog) were followed by the highest number of collisions (Louw et al., 2016). These results illustrate, therefore, that scenarios that encourage driver gaze towards the road centre are likely to bring drivers back into the loop more efficiently by facilitating better situation awareness/hazard perception during the transfer of control from highly automated driving. Although further work is required to validate this proposal, these findings suggest that any information presented to drivers during automation should be placed near the centre of the road, akin to a Head Up Display. Clearly, the interaction between presenting such information towards the road centre during automation, and the consequent effect on driver distraction, needs further investigation.

An encouraging finding from these studies was that, regardless of screen manipulation, drivers' understanding of the automated system and uncertainty events increased as time progressed. This was illustrated by observations in vertical gaze dispersion, which was significantly reduced after the first event. There was also an increase in horizontal dispersion of gaze upon approach to the second critical event in automation; suggesting drivers prepared themselves, for instance by looking towards the adjacent lane before a lane change, to avoid collision with the lead vehicle.

It is important to note that the manipulations used in this study do not provide a complete assessment of the OOTL state. As disengagement from the driving task was involuntary and experimenter-induced, the effects are likely underestimated. Under normal automated driving conditions, drivers' withdrawal of attention, and therefore disengagement from feedback of driving relevant information, is generally self-induced and voluntary. Maintaining consistent and voluntary disengagement from all aspects of the driving task in a controlled setting highlights a key challenge in attempting to investigate the OOTL state. Moreover, gaze dispersion is only one of several measures that should be examined to investigate whether automation unduly impedes drivers' abilities to regain full cognitive and physical control. Therefore, natural progressions of this work are to analyse how drivers' visually process road hazards in critical takeover scenarios, and to establish how well calibrated drivers' vehicle control is to the criticality of an unfolding scenario, following a takeover. Future studies should also consider the effect of longer periods of placing drivers OOTL on such measures, as well as HMI solutions that provide informative, yet non-intrusive system feedback.

Acknowledgements

This research was conducted as part of the AdaptIVe project, co-funded by the European Union under the 7th Framework Programme, grant agreement number 610428. The authors would like to thank Oliver Carsten, Georgios Kountouriotis, Ruth Madigan, Gustav Markkula, and the two anonymous reviewers for their helpful comments and discussions, and Michael Daly and Andrew Tomlinson for their assistance in implementing the simulator scenarios.

References

Anderson, J. M., Kalra, N., Stanley, K. D., Sorensen, P., Samaras, C., Oluwatola, O. A., 2014. Autonomous Vehicle Technology: A Guide for Policymakers. Publication RR-443-RC. RAND Corporation, Santa Monica, CA.

Bainbridge, L., 1983. Ironies of automation. Automatica, 19(6), pp.775-779.

Beller, J., Heesen, M., Vollrath, M., 2013. Improving the Driver–Automation Interaction: An Approach Using Automation Uncertainty. Hum Factors 55 (6), 1130-1141. DOI: 10.1177/0018720813482327

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., Merat, N., 2012. Control task substitution in semi-automated driving: Does it matter what aspects are automated? Hum Factors 54, 747–761. DOI: 10.1177/0018720812460246

Damböck, D., Weißgerber, T., Kienle, M., Bengler, K., 2013. Requirements for cooperative vehicle guidance. Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems, The Hague, The Netherlands (pp. 1656–1661). DOI: 10.1109/ITSC.2013.6728467.

De Winter, J. C. F., Happee, R., Martens, M. H., Stanton, N. A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. Transportation Res F-TRAF 27, 196-217. DOI: 10.1016/J.TRF.2014.06.016

Endsley, M. R., 2006. Situation awareness. In: Salvendy, G. (Ed.), Handbook of human factors and ergonomics, Wiley, New York, pp. 528–542. DOI:10.1002/0470048204.ch20

Endsley, M. R., Kiris, E. O., 1995. The out-of-the-loop performance problem and level of control in automation. Hum Factors 37 (2), 381-394. DOI: 10.1518/001872095779064555

Endsley, M. R., Jones, D. G., 2004. Designing for situation awareness: An approach to user centered design, Taylor & Francis, New York. DOI: 10.1177/106480460401200310

Engström, J., Johansson, E., Ostlund, J., 2005. Effects of visual and cognitive load in real and simulated motorway driving. Transportation Res F-TRAF 8, 97-120.

Field, A. P., 2009. Discovering statistics using SPSS: And sex and drugs and rock 'n' roll (3rd ed.). Sage, London.

Foley, J.P., 2009. Now you see it, now you don't: visual occlusion as a surrogate distraction measurement technique. In: Young, K., Lee, J.D., Regan., M.A. (Eds.), Driver distraction: theory, effects and mitigation, CRC Press, Boca Raton, pp.123-134.

Gartenberg, D., Breslow, L. J., McCurry, M., Trafton, G. J., 2013. Situation awareness recovery. Hum Factors 56, 710–727. DOI:10.1177/0018720813506223

Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. “Take over!” How long does it take to get the driver back into the loop? Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57, 1938-1942.

Gugerty, L., 2011. Situation awareness in driving. In: Lee, J., Rizzo, M., Fischer, D., Caird, J. (Eds.), Handbook for driving simulation in engineering, medicine and psychology, CRC Press, Boca Roca, FL. DOI:10.1201/b10836-20.

Helldin, T., Falkman, G., Riveiro, M., Davidsson, S., 2013. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. DOI: 10.1145/2516540.2516554

Horswill, M. S., McKenna, F. P., 2004. Drivers’ hazard perception ability: Situation awareness on the road. In: Banbury, S., Tremblay, S., (Eds.), A Cognitive Approach to Situation Awareness: Theory and Application, Ashgate, Hampshire, England, pp. 155-175. DOI: 10.1109/cogsima.2013.6523837

IBM Corp., 2012. IBM SPSS statistics (Version 21.0) [Computer software]. Armonk, NY: IBM Corp.

Infinity Motors, 2015. 2015 Infiniti Q50 Press Kit. Retrieved 28 November, 2015. Retrieved from http://infinitinews.com/en-US/infiniti/usa/presskits/us-2015-infiniti-q50-press-kit

Kienle, M., Damböck, D., Kelsch, J., Flemisch, F. & Bengler, K., 2009. Towards an H-Mode for highly automated vehicles: driving with side sticks. Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009), Sep 21-22 2009, Essen, Germany, p. 19-23.

Körber, M., Cingel, A., Zimmermann, M., Bengler, K., 2015. Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. Procedia Manufacturing 3, 2403–2409.

Lerner, N., Baldwin, C., Stephen Higgins, J., Lee, J.D., Schrooler, J., 2015. Mind wandering while driving: What does it mean and what do we do about it? Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting, pp. 1686-1690. DOI: 10.1177/1541931215591364

Louw, T., Kountouriotis, G., Carsten, O., Merat, N., 2015. Driver Inattention During Vehicle Automation: How Does Driver Engagement Affect Resumption of Control? Proceedings of the 4th International Conference on Driver Distraction and Inattention. DOI: 10.13140/RG.2.1.2017.0089

Louw, T., Merat, N., Jamson, A. H., 2015. Engaging with Highly Automated Driving: To be or not to be in the loop? Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 190-196. Snowbird, Utah. DOI:10.13140/RG.2.1.2788.9760

Louw, T., Madigan, R., Carsten, O., Merat, N., 2016. Were they in the loop during automated driving? Links between visual attention and crash potential. Inj Prev. DOI: 10.1136/injuryprev-2016-042155

Merat, N., Jamson, A. H., 2008. How do drivers behave in a highly automated car? Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 514–521. DOI:10.1016/J.TRC.2013.02.008

Merat, N., Jamson, A. H., Lai, F. C., Carsten, O., 2012. Highly automated driving, secondary task performance, and driver state. Hum Factors 54, 762–771. DOI: 10.1177/0018720812442087.

Merat, N., Jamson, A. H., Lai, F., Daly, M., Carsten, O. M., 2014. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. Transportation Res F-TRAF 26, 1–9. DOI:10.1016/J.TRF.2014.05.006.

Neter, J., Wasserman, W., Kutner, M.H., 1990. Applied Linear Statistical Models, Irwin, Homewood, IL. DOI:10.1080/00401706.1990.10484703

Norman, S.D. and Orlady, H.W., 1989. Flight deck automation: Promises and realities. Moffett Field, California: National Aviation and Space Administration.

Norman, D.A., 1990. The problem with automation: inappropriate feedback and interaction, not'over-automation'. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 327(1241), pp.585-593.

Pick, A.J., Cole, D.J., 2006. Neuromuscular dynamics in the driver-vehicle system. Vehicle Syst Dyn 44, 624–631. DOI: 10.1080/00423110600882704

Posner, M.I., 1980. Orienting of attention. Q J Exp Psychol 32, 3-25. DOI: 10.1080/00335558008248231

Rasmussen, J., Rouse, W.B., 2013. Human detection and diagnosis of system failures (Vol. 15). Springer Science & Business Media.

Rosenbloom, S., 2012. The Travel and Mobility Needs of Older People Now and in the Future. In: Coughlin, J.F, D'Ambrosio, L.A. (Eds.), Aging America and Transportation: Personal Choices and Public Policy, Springer, New York, pp. 39–56. DOI: 10.1080/01634372.2013.796221

Rudin-Brown, C.M., Parker, H.A., 2004. Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. Transportation Res F-TRAF 7 (2), 59–76.

Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, C.W., Ward, J.L., 1967. The attentional demand of automobile driving. Highway research record, (195).

Society of Automotive Engineers [SAE] (2014. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (Standard No. J3016). SAE International. Retrieved from: http://standards.sae.org/j3016_201401/

Stanton, N. A., Young, M. S., 1998. Vehicle automation and driving performance. Ergonomics 41 (7), 1014-1028. DOI: 10.1080/001401398186568

Strand, N., Nilsson, J., Karlsson, I.M. and Nilsson, L., 2014. Semi-automated versus highly automated driving in critical situations caused by automation failures. Transportation Res F-TRAF 27, 218-228. DOI: 10.1016/J.TRF.2014.04.005

Tabachnick, B. G., Fidell, L. S., 2007. Using multivariate statistics, Pearson/Allyn & Bacon, Boston, MA.

Tesla Motors, 2015. Your Autopilot has arrived. Retrieved 28 November, 2015, from <https://www.teslamotors.com/blog/your-autopilot-has-arrived>

Volvo Cars, 2015. This is autopilot. Retrieved 28 November, 2015, from <http://www.volvocars.com/intl/about/our-innovation-brands/intellisafe/intellisafe-autopilot/this-is-autopilot/semi-autonomous-tech>

Wang, Y., Reimer, B., Dobres, J., Mehler, B., 2014. The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. Transportation Res F-TRAF 26, 227-237. DOI:10.1016/J.TRF.2014.08.003

Wiener, E.L., Curry, R.E., 1980. Flight-deck automation: Promises and problems. Ergonomics, 23(10), pp. 995-1011.

Zeeb, K., Buchner, A., Schrauf, M., 2015. What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident Anal Prev 78, 212–221.