The notorious B.I.T: The effects of a ransomware and a screen failure on distraction in automated driving

William Payre *, Jaume Perelló-March, Ashwini Kanakapura Sriranga, Stewart Birrell

National Transport Design Centre, Coventry University, Coventry CV1 2TT, UK

ARTICLE INFO

Keywords: Eye-tracking Cybersecurity Automated driving Failure Distraction Ransomware

ABSTRACT

Connected and automated vehicles are vulnerable to cyber-attacks, which may jeopardise their safe and efficient operation and, as a result, negatively affect drivers’ behaviour. A major concern for such cyber-attacks is visual distraction inside the vehicle, which is one of the main causes of road accidents. In this empirical research using a driving simulator, 38 participants drove in a conditionally automated vehicle and experienced two types of failure: explicit (i.e., ransomware attack appearing on the in-vehicle screen) and silent (i.e., turn signals failed to activate on the in-vehicle screen and instrument cluster), while engaged in a non-driving related task. Drivers’ gaze behaviour, in terms of number and duration of fixation, were collected and analysed. Results showed that the HMI where the ransomware was displayed was the area of interest drivers looked at the most. The majority of drivers failed to notice that the turn signal was faulty. Nearly half of drivers looked at the ransomware for more than 12 s while driving. No effect on the timing of failure on gaze behaviour was observed. This research evidenced that ransomware attacks are distractive and pose significant risks to road safety – with one participant crashing the vehicle after resuming manual control. Data also evidenced that such connected vehicles are unlikely to meet NHTSA’s distraction guidelines for safe use of in-vehicle devices.

1. Introduction

The development of advanced infrastructure to support Intelligent Transportation Systems (ITS) and vehicle to everything (V2X) communications, along with the mass adoptions of connected smartphones, has facilitated the rapid rise of the connected and automated vehicles (CAV). It is estimated that the connected car market will be worth circa $192 billion (U.S. dollars) by 2028 (Fortune Business Insights, 2021). While CAVs offer a range of benefits in terms of safety (e.g. lane keeping assist and adaptive cruise control) and comfort (e.g. dynamic navigation and handing-over the driving task), they also posit challenges in terms of cybersecurity (ISO/SAE 21434, 2021). Indeed, with the integration of drive-by-wire systems within cars, CAVs at moderate and high levels of automation (e.g. levels 2, 3 and 4; SAE J3016, 2020) become vulnerable to software and hardware failures, which bears implications for road safety (Dede et al., 2021; ISO/TR 4804, 2020). For instance, in January 2021, the National Highway Traffic Safety Administration (NHTSA; Ridella, 2021) issued a recall to the car manufacturer Tesla for about 158,000 Model S and X vehicles because of touchscreen failures, where the turn signals functionality was faulty. The in-vehicle system did not warn drivers of this failure, and the turn signal did not activate when it needed to while in automated driving mode. This type of malfunction is referred to as a silent failure, whereby the system fails to notify individuals of its incapacity to operate reliably.

* Corresponding author at: National Transport Design Centre, Coventry University, Swift road, Coventry CV1 2TT, UK.
E-mail address: william.payre@coventry.ac.uk (W. Payre).
SAE level 2 automated cars partially relieve individuals from the driving task. Silent failures in these automated vehicles lead to decreased engagement in monitoring the road, poorer take-over performance (Alambeigi et al., 2021; Louw et al., 2019; Mole et al., 2020) and delayed brake response times (Bianchi-Piccinini et al., 2020). Explicit failures, as opposed to silent ones, are notified to the users. With respect to CAV, drivers can be warned of a system failure via the in-vehicle display or an acoustic warning. Previous research examined the effect of automation limitations and explicit failures on driving behaviour, attitudes and performance (e.g. Eriksson and Stanton, 2017; Payre et al., 2017; Zhang et al., 2019). However, little is known on how an external threat, such as hacking the computerised vehicle system, affects CAV drivers. This is surprising as recent evidence suggests that safety, privacy and data security are the main public concerns regarding CAV in the USA (Lee & Hess, 2022). Furthermore, conditionally automated (SAE level 3, such as Mercedes-Benz Drive Pilot) and connected cars have been temporarily allowed on UK roads, although no such vehicles are listed for use at the time of writing (Centre for Connected and Autonomous Vehicles, 2022). Hence, this empirical research aims to shed light on the previously unexplored effect of silent and explicit failures on drivers’ gaze behaviour in a conditionally automated car (i.e. SAE level 3), where driver’s supervision is limited and not mandatory, resulting in more engagement in non-driving related tasks (NDRT; Cunningham & Regan, 2018). More specifically, we examine how different failures occurring during potentially hazardous situation (i.e. overtaking manoeuvres and merging sections) and their timing (i.e. early or late failures) affect drivers in terms of gaze behaviour and take-over of automation. The focus is on driver’s gaze behaviour as driving mostly is a visual task and driver distraction is a prominent cause of crashes (Hanowski et al., 2017; Treat et al., 1979). Distraction occurs when individuals’ attention is diverted away from the driving task by another object or activity (Horberry et al., 2006; Young & Salmon, 2012). For instance, distraction comes to the fore when drivers look at in-vehicle screens instead of the road, leading to driving errors stemming from high visual attention demand (Young & Regan, 2007). The silent failure investigated in this research imitates the 2021 Tesla recall where turn signals failed to activate. The explicit failure consists in a ransomware threatening drivers’ privacy and encrypting personal data until making a payment in cryptocurrency. The research question is “Are silent and explicit in-vehicle screen failures in CAV an emerging source of distraction for drivers?”.

1.1. Aims and hypotheses

The present research will further our understanding of CAV failures in terms of gaze and take-over behaviour when exposed to a cyber-attack (i.e., ransomware) and a screen failure (i.e., no turn signals) at two different timings (i.e., early vs late failure). The type of failure is an exploratory variable, whereas the early failures should affect drivers’ trust and monitoring behaviour (Lee & See, 2004). Gaze behaviour is analysed during the entire journey and after each type of failure. This will allow comprehending whether drivers allocate their attention to the in-vehicle screen if they resume manual control, which could be detrimental to road safety as they look away from the road.

We hypothesised that:

• the number (a) and the duration (b) of fixations on the in-vehicle screen are greater after the explicit failure than after the silent failure and when there is no failure (H1);
• after the explicit failure, the number (a) and the duration (b) of fixations on the in-vehicle screen are lower when drivers have resumed control of the vehicle than when they have not (H2);
• the number (a) and the duration (b) of fixations on the in-vehicle screen are higher when the failures occur early during the scenario than late (H3);

2. Material and methods

2.1. Driving simulator

The experiment was carried out using a moving-base high-fidelity driving simulator equipped with a full-body Ford Focus and advanced Unity graphics (Fig. 1a and 1b). The main visuals were provided by five projectors with a 1920 x 1200 px display resolution.

![Fig. 1. a (left) The driving simulator, the curved screen and the monitor displaying the rear-view mirror. 2b (right) view of the Unity based driving environment.](image-url)
at 60 Hz, rendered on a 4.75 m, 270° curved screen. Road motion was generated through a hydraulic system with three degrees of freedom coupled with realistic environmental sound administered in stereo via 2 × 20 W speakers. The simulated driving automation was capable of lane-keeping, adaptive cruise control, emergency braking, and overtaking slower vehicles. As the simulator is fully driver-in-the-loop, the driver could initiate automated driving mode, or regain manual control from automated driving at any point in the driving scenario (rather than only at predefined scripted points). Automated mode was engaged by tapping a button on the in-vehicle interface (Fig. 2b), the driver could take over manual control of the driving task by either tapping the same button, or by applying force to the steering wheel or brake pedal. This level of control offered by the simulator allows for a more naturalistic assessment of driving behaviour.

2.2. Human-machine interface (HMI)

A Raspberry Pi 3 with a 7′′ resistive touchscreen display was mounted on top of the central console accommodating an infotainment interface and an in-house python app to communicate with the driving simulator. The information displayed included a static map view on the left side, and a dynamic vehicle status view. Vehicle status included the mode of control (i.e., manual or automated) and the turn signals (green arrows) (Fig. 2).

2.3. Experimental design

This experiment entailed two experimental conditions and a control condition (see Figs. 3, 4, 5 & 6). A mixed within-between-subjects design with repeated measures was adopted, with the type of failure as a within-subject variable (silent vs explicit) and the timing of failure as a between-subjects variable (early vs late). Assignment to conditions was counterbalanced. Driving scenarios took place in a digital twin (i.e. a 3D virtual replica) of Coventry, UK, which included 15 miles of suburban roads and motorway.

All participants first completed a familiarisation trial consisting in driving manually for at least 5 min. Participants were asked to follow the predefined navigation instructions, comply with the UK Highway Code and drive as they usually do. They were presented a description of the conditionally automated driving (SAE-L3) capabilities of the vehicle. These included automated longitudinal and lateral control of the car, object recognition (i.e., road signals, traffic lights and car indicators), and overtaking manoeuvres. After reaching a motorway in the familiarisation trial, a visual and audio message was presented suggesting participants to activate automated driving by tapping a steering wheel icon on the HMI (Fig. 2). After 2 min of automated driving, the vehicle safely pulled-over and stopped.

Each condition (control, silent and explicit) lasted for approximately 12 min (Fig. 3). Participants were prompted to engage in the non-driving related task (NDRT, i.e., word search) as soon as they activated automated driving. Word search is a visually demanding task encouraging drivers to look away from the road. The goal was to find as many words as possible before the end of the trial. However, they were explicitly allowed to resume manual control and reengage automated driving at any time they wished.

In the control condition, participants started driving manually at a roundabout leading to the motorway (A45). Approximately 30 s after merging-in, automated driving was made available and participants were prompted to engage it as in the familiarisation trial. This condition did not include any silent or explicit failures. However, the following event happened twice during the control condition: the participant’s car, which was in automated mode, slowed down, then overtook a larger and slower vehicle (i.e., van) merging into

Fig. 2. The HMI touchscreen display in the vehicle.
The turn signals were activated during the manoeuvre. The silent condition involved the same procedure as in the control condition, although including a silent failure consisting of the turn signals failing to activate while overtaking a van merging into the motorway from the slip road (Fig. 5). As a result, the turn signal was displayed neither on the HMI nor on the instrument cluster. One group experienced an early silent failure (~2 min after automated driving was engaged), whilst the other experienced a late silent failure (~8 min after automated driving was engaged).
The explicit condition used a procedure similar to the control condition, but with an explicit HMI failure (Fig. 6). It consisted of a cyber-attack, more specifically a ransomware attack, intruding into the vehicle’s computerised system to access stored information. More specifically, the ransomware encrypted the personal information entered by the participant on the HMI at the beginning of the experiment. A full-screen ransomware image, adapted from Wolf and Lambert (2017) and the 2017 WannaCry ransomware, was displayed on the in-vehicle screen with the name of the participant. A £200 payment worth of Bitcoin was required to protect the personal data. A “pay after my trip” button enabled participants to navigate back to the original screen (Fig. 7a and 7b). The ransomware was displayed at the start during an overtaking manoeuvre similar to those performed in the control and silent conditions. All road vehicles’ appearance changed across conditions to avoid priming participants. The merging-in vehicles used during the overtaking manoeuvres were always vans, but their model and colour changed.

Participants were assigned to either the early (failure happened ~ 2 min after activation of automated driving) or late (failure happened ~ 8 min after activation of automated driving) timing conditions depending on their age and sex to balance across groups. The experimenter calibrated the eye-tracking system and asked participants to login their personal details into the infotainment screen: name, surname, email and password. These details would later be used as leverage in the ransomware attack.

This procedure was repeated for the three experimental conditions, which were counterbalanced to control for any order effects. Before, during and after the experiment, each participant completed a set of scales and answered open-ended questions. These results are not reported in this paper.

2.4. Participants

Thirty-eight participants were recruited to participate in the study. The sample consisted of 16 females and 18 males, with an average age of 36.2 (SD = 12.5), 0 to 43 years of driving experience (M = 15.7, SD = 13.1) and a yearly mileage of 7737 miles (min = 0, max = 20000, SD = 5891). Participants voluntarily agreed to take part in this experiment and were free to withdraw at any point. All of them held a valid driving license, were aged 18 or above and had normal or corrected to normal vision. After the experiment, they received a £20 voucher. Recruitment and data collection methods received approval from Coventry University ethics committee.

2.5. Measures

Eye-tracking data was gathered using a fixed-base, three cameras Smart Eye system (Smart Eye AB, Gothenburg, Sweden). Cameras were fixed on the left, centre and right sides of the dashboard to ensure gaze tracking was optimal (see Fig. 2). Data was sampled at 60 Hz with Smart Eye Pro 7.2 software. The number of fixations and their duration were the features extracted. Number of fixations refers to the total number of fixations identified on the defined areas of interest (i.e., HMI, instrument cluster and right wing mirror). It measures the importance of a display in terms of visual attention (Pitts et al., 1950). Duration of fixations is defined as the period in seconds during which the fixations are measured in single area of interest. It refers to total fixation time, not to the mean single glance duration. It measures how difficult it is to extract or interpret information on a display (Hergeth et al., 2016).

In addition, the number of fixations greater than 2 s and the number of participants who looked at the HMI more than 12 s after each failures were collected to check against the NHTSA acceptance criteria to perform secondary tasks while driving (NHTSA, 2013). In the present experiment, we considered that looking at the HMI after the explicit and silent failures was an exploratory task where drivers could look at the HMI and interact by tapping buttons while driving. We also extracted fixation count and duration for 3 min after each failure. A fixation rejection threshold was set at below 60 ms as 50–75 ms is the shortest duration range (Rayner, 2009). In addition, we conducted semi-structured interviews after the experiment to gather participants’ impressions. These declarative data has been published in a parallel paper (for more details see Payre et al., 2022).

2.6. Data analysis

Gaze behaviour data (i.e., count and duration) were analysed within conditions (i.e., explicit, silent and control) and within areas of interest (i.e., HMI, instrument cluster and wing mirror). Since the assumption of normality was violated for the number and duration of fixations (including data collected after each failure/overtaking event), we ran the following non-parametric tests:

- Related-Samples Friedman’s Two-Way Analysis of Variance, to compare the effect of the three within-subject conditions (i.e. control, silent failure and explicit failure) on gaze behaviour. Post-hoc pairwise comparisons of the independent samples were conducted using Wilcoxon signed-rank tests. We also applied a Bonferroni correction at p < 0.017 to protect from Type 1 Error.
- Mann-Whitney U tests to investigate the effect of the type of vehicle control (manual vs automated) and the timing of the failure (early vs late) on gaze behaviour.

3. Results

One participant was excluded from the eye-tracking data analysis due to missing data. A second participant who suffered from simulator sickness was not included in the analysis after dropping out from the study.
3.1. Descriptive statistics

Descriptive statistics for the gaze behaviour on the different areas of interest are presented in Table 1.

Gaze behaviour on the HMI with respect to the NHTSA driver distraction criteria are presented in Fig. 8. In general, and regardless whether they resumed manual control, almost 40 % of the participants (N = 14) spent more than 12 s looking at the HMI after the explicit failure (i.e., ransomware). Among those 14 participants, 9 (64 %) of them had at least one fixation longer than 2 s at the ransomware. One participant looked at the HMI for more than 2 s at once after the silent failure (i.e., no turn signals), and none in the control condition. No participants looked at the HMI for more than 12 s after the silent failure and the similar corresponding event in the control condition. These descriptive statistics aim to provide an overview on gaze behaviour towards the HMI across conditions.

3.2. Effect of the type of failure

The number (H1a) and the duration (H1b) of fixations on the in-vehicle screen (i.e., HMI) are greater after the explicit failure than after the silent failure and when there is no failure.

A statistically significant difference in the number of fixations on the HMI between conditions was found (Fig. 9), $\chi^2(2) = 36.014, p < 0.001$. Post-hoc analysis revealed that the total duration of fixations on the HMI during the explicit failure condition (Mdn = 19) was greater than the silent failure (Mdn = 5, Z = -4.174, p < 0.001) and control (Mdn = 4, Z = -4.707, p < 0.001) conditions. $H_{1a}$ is accepted. A further exploration of these results showed that in the explicit failure condition, the number of fixations varied across the areas of interest $\chi^2(2) = 28.222, p < 0.001$. Post-hoc analysis revealed that the number of fixations on the HMI (Mdn = 19, Z = -4.149, p < 0.001) and on the IC (Mdn = 15, Z = -4.062, p < 0.001) were greater than on the mirror (Mdn = 2).

A statistically significant difference in the number of fixations on the IC between conditions was found, $\chi^2(2) = 7.036, p = 0.030$. Post-hoc analysis revealed that the number of fixations on the IC in the explicit failure (Mdn = 15, Z = -2.236, p = 0.025) and the silent failure (Mdn = 15, Z = -2.070, p = 0.038) conditions were greater than the control condition (Mdn = 12). A further exploration of these results showed that during the silent failure, the number of fixations varied across the areas of interest $\chi^2(2) = 28.884, p < 0.001$. Post-hoc analysis revealed that during the silent failure, the number of fixations on the IC (Mdn = 15) was greater than the HMI (Mdn = 5, Z = -4.105, p < 0.001) and the mirror (Mdn = 3, Z = -4.293, p < 0.001).

A statistically significant difference in the duration of fixations on the HMI between conditions was found (Fig. 9), $\chi^2(2) = 35.549, p < 0.001$. Post-hoc analysis revealed that the total duration of fixations on the HMI during the explicit failure condition (Mdn = 12.864) was greater than during the silent failure (Mdn = 2.192, Z = -6.450, p < 0.001) and control conditions (Mdn = 1.854, Z = -4.823, p < 0.001). $H_{1b}$ is accepted. A further exploration of these results showed that during the explicit failure condition, the duration of fixations varied across the areas of interest $\chi^2(2) = 29.556, p < 0.001$. Post-hoc analysis revealed that the duration of fixations on the HMI (Mdn = 12.864, Z = -4.627, p < 0.001) and on the IC (Mdn = 6.368, Z = -3.990, p < 0.001) were greater than on the mirror (Mdn = 1.136).

A statistically significant difference in the duration of fixations on the IC between conditions was found, $\chi^2(2) = 7.056, p = 0.029$. Post-hoc analysis revealed that the duration of fixations on the IC in the explicit failure condition (Mdn = 6.368) was greater than the control condition (Mdn = 5.104, Z = -2.199, p = 0.028). A further exploration of these results showed that during the silent failure condition, the duration of fixations varied across the areas of interest $\chi^2(2) = 24.182, p < 0.001$. Post-hoc analysis showed that the duration of fixations on the IC (Mdn = 6.624) was greater than the HMI (Mdn = 2.192, Z = -3.975, p < 0.001) and the mirror (Mdn = 1.184, Z = -4.210, p < 0.001).

3.3. After the failure/event

A statistically significant difference in the number of fixations on the HMI after the event between conditions was found (Fig. 9), $\chi^2(2) = 4.210, p = 0.038$. Further exploration revealed that the number of fixations varied across the areas of interest $\chi^2(2) = 6.624, p = 0.039$. In the explicit failure condition (Mdn = 15), $\chi^2(2) = 0.025$. A further exploration of these results showed that during the explicit failure condition, the number of fixations varied across the areas of interest $\chi^2(2) = 28.222, p < 0.001$. Post-hoc analysis revealed that the number of fixations on the HMI (Mdn = 19, Z = -4.149, p < 0.001) and on the IC (Mdn = 15, Z = -4.062, p < 0.001) were greater than on the mirror (Mdn = 2).

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count (number)</th>
<th>Duration (seconds)</th>
<th>Count After Failure (number)</th>
<th>Duration After Failure (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Explicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wing Mirror</td>
<td>6.595</td>
<td>9.509</td>
<td>3.358</td>
<td>5.523</td>
</tr>
<tr>
<td>Silent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMI</td>
<td>7.081</td>
<td>7.907</td>
<td>3.613</td>
<td>4.562</td>
</tr>
<tr>
<td>Wing Mirror</td>
<td>6.405</td>
<td>11.019</td>
<td>3.195</td>
<td>5.552</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMI</td>
<td>6.595</td>
<td>7.635</td>
<td>3.380</td>
<td>4.207</td>
</tr>
<tr>
<td>Wing Mirror</td>
<td>5.513</td>
<td>8.112</td>
<td>2.726</td>
<td>4.232</td>
</tr>
</tbody>
</table>
Transportation Research Part F: Psychology and Behaviour 94 (2023) 42–52

(2) = 58.274, \( p < 0.001 \). Post-hoc analysis revealed that the number of fixations on the HMI after the explicit failure (Mdn = 16) was greater than after the silent failure (Mdn = 1, \( Z = -5.088, p < 0.001 \)) and control (Mdn = 1, \( Z = -5.162, p < 0.001 \)). A further exploration of these results showed that after the explicit failure, the number of fixations varied across the areas of interest \( \chi^2(2) = 42.606, p < 0.001 \). Post-hoc analysis revealed that the number of fixations on the HMI (Mdn = 16) was greater than on the IC (Mdn = 5, \( Z = -4.005, p < 0.001 \)) and on the mirror (Mdn = 1, \( Z = -5.096, p < 0.001 \)). In addition, the number of fixations on the IC was greater than on the mirror (\( Z = -3.833, p < 0.001 \)).

There was no effect of the conditions for the number of fixations on the IC (\( p = 0.081 \)).

A statistically significant difference in the duration of fixations after the event on the HMI between conditions was found (Fig. 9), \( \chi^2(2) = 56.168, p < 0.001 \). Post-hoc analysis revealed that the duration of fixations on the HMI after the explicit failure (Mdn = 9.376) was greater than after the silent failure (Mdn = 0.336, \( Z = -5.159, p < 0.001 \)) and control (Mdn = 0.224, \( Z = -5.159, p < 0.001 \)). A further exploration of these results showed that after the explicit failure, the duration of fixations varied across the areas of interest \( \chi^2(2) = 41.378, p < 0.001 \). Post-hoc analysis revealed that the duration of fixations on the HMI (Mdn = 9.376, \( Z = -5.159, p < 0.001 \)) and on the IC (Mdn = 1.808, \( Z = -3.363, p = 0.001 \)) were greater than on the mirror (Mdn = 0.224). In addition, the duration of fixations on the IC was greater than on the mirror (\( Z = -3.833, p < 0.001 \)).

3.4. Effect of the type of vehicle control (manual vs automated)

After the explicit failure, the number (\( H_{2a} \)) and the duration (\( H_{2b} \)) of fixations on the in-vehicle screen (i.e., HMI) are lower when drivers have resumed control of the vehicle than when they have not.

An Independent-Samples Mann-Whitney U Test did not reveal any statistically significant effect of the type of vehicle control on the
number (p = 1.000) or the duration of fixations (p = 0.871). An exploration of the descriptive statistics showed that 4 out of 9 participants who resumed manual control after the explicit failure spent more than 12 s on the HMI. Whereas only 8 out of 23 who did not resume control after the explicit failure spent more than 12 s on the HMI. Finally, 2 out of 9 drivers who resumed control after the explicit failure looked at the HMI for more than 2 s at once. \( H_{3a} \) and \( H_{3b} \) are rejected.

In addition, one participant crashed the vehicle after resuming control and going through the hard shoulder.

3.5. Effect of the timing of failure

The number (\( H_{3a} \)) and the duration (\( H_{3b} \)) of fixations on the in-vehicle screen (i.e., HMI) are higher when the failures/events occur early during the scenario than late.

A Mann-Whitney U test was performed to test for the effect of the timing of failure (i.e., early vs late) on the number and duration of fixations on the HMI. No statistically significant effect of the timing of failure on gaze behaviour was observed for the explicit (\( p = 0.358 \)) and silent (\( p = 0.940, p = 0.730 \)) conditions. and control (\( p = 0.221, p = 0.178 \)) conditions.

3.6. Summary of the statistical results and key findings

Key findings (Table 2):

- Explicit (ransomware attack) was the type of failure that drew drivers’ attention the most
- Glances toward the HMI were similarly distributed between automated and manual driving
- No significant effects for the timing of failure were observed on gaze behaviour

4. Discussion

4.1. Explicit vs Silent failure

We first hypothesised that the number and the duration of fixations on the in-vehicle screen would be greater after the explicit failure than after the silent failure and when there was no failure (\( H_1 \)). This hypothesis was partially accepted. As expected, the explicit failure was more conspicuous than the silent one, as shown by the increased number of fixations and longer duration. This new empirical finding is in line with recent evidence suggesting that drivers allocate some attention towards a HMI when provided with an explanation for the occurrence of a failure (Kraft et al., 2020; Ulahannan et al., 2021).

Concerning the silent failure condition, the findings indicated a greater number and longer duration of fixations on the instrument

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count (number)</th>
<th>Duration (seconds)</th>
<th>Count After Failure (number)</th>
<th>Duration After Failure (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of Interest</td>
<td>Test Statistic</td>
<td>p value</td>
<td>Test Statistic</td>
<td>p value</td>
</tr>
<tr>
<td>HMI</td>
<td>( \chi^2(2) = 36.014 )</td>
<td>&lt; 0.001</td>
<td>( \chi^2(2) = 35.549 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Explicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent</td>
<td>( Z = -4.174 )</td>
<td>&lt; 0.001</td>
<td>( Z = -4.650 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Control</td>
<td>( Z = -4.707 )</td>
<td>&lt; 0.001</td>
<td>( Z = -4.823 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Inst. Cluster</td>
<td>( \chi^2(2) = 7.036 )</td>
<td>0.030</td>
<td>( \chi^2(2) = 7.056 )</td>
<td>0.029</td>
</tr>
<tr>
<td>Explicit</td>
<td>( Z = -2.236 )</td>
<td>0.025</td>
<td>( Z = -2.199 )</td>
<td>0.028</td>
</tr>
<tr>
<td>Silent</td>
<td>( Z = -2.070 )</td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wing Mirror</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit</td>
<td>( \chi^2(2) = 28.222 )</td>
<td>&lt; 0.001</td>
<td>( \chi^2(2) = 29.556 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Silent</td>
<td>( Z = -4.149 )</td>
<td>&lt; 0.001</td>
<td>( Z = -4.627 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Control</td>
<td>( Z = -4.062 )</td>
<td>&lt; 0.001</td>
<td>( Z = -3.990 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Inst. Cluster</td>
<td>( \chi^2(2) = 28.884 )</td>
<td>&lt; 0.001</td>
<td>( \chi^2(2) = 24.182 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Wing Mirror</td>
<td>( Z = -4.105 )</td>
<td>&lt; 0.001</td>
<td>( Z = -3.975 )</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Control</td>
<td>( Z = -4.293 )</td>
<td>&lt; 0.001</td>
<td>( Z = -4.210 )</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
cluster (i.e., where the speedometer and other vehicle information, including the indicators, are traditionally presented) than the HMI and wing mirror. This would suggest that drivers remained vigilant during this scenario despite performing the NDRT. Notwithstanding, this acute vigilance toward the instrument cluster was not statistically significant during the 3 min following the silent failure, suggesting that drivers tried to stay vigilant throughout the entire trial, yet most of them missed the silent failure. This is congruent with the low number of participants (6 out of 36) who resumed control after the silent failure, probably because they did not notice it. This is problematic with respect to road safety (Cunningham & Regan, 2018), compliance with the Highway Code in the UK (i.e., turn signals were mandatory in this instance) and SAE guidelines (i.e., drivers must be aware of their surroundings and pay attention to the status of the system). Actually, drivers complied with SAE guidelines as they paid attention to their surroundings and regularly checked the HMI, the instrument cluster and the wing mirror in all three conditions, but most of them could not detect a subtle failure of the automated driving system. They seemed to be on the loop – i.e., not in physical control of the vehicle, but monitoring the driving situation (Merat et al., 2019). Therefore, this research poses the important question of should we consider drivers at fault because they failed to notice a silent failure, although paying attention to the status of the system? This an open question, but there is abundant literature referring to the mediocre ability of individuals to remain vigilant over extended periods of time in automated human–machine systems (see Warm, Matthews & Finomore Jr, 2018 for a review). Despite the turn signals not being displayed on both the instrument cluster and the HMI, drivers failed to notice they did not work during an automated overtaking manoeuvre while being engaged in a NDRT. This result is pertinent from an insurance perspective to determine who is responsible in that situation: the original equipment manufacturer (OEM) because of the faulty device or the driver for not picking up and correcting the silent failure?

This finding on glance behaviour is consistent with previous literature showing the influence of secondary task engagement on drivers’ situation awareness during silent failures (Louw et al., 2019), but also suggests that participants had low to moderate levels of trust in the CAV. Indeed, the number of fixations is considered “a more direct measure of automation trust and reliance than fixation durations” (Hergeth et al., 2016). Therefore, although drivers in this research seemed to have moderate levels of trust in the system and tried to remain in the loop, they still missed the silent failure.

The specificity of this experiment is that drivers drove in a conditionally automated vehicle (SAE L3) and were conducting a NDRT, whereas previous research investigated the effect of silent failures with SAE L2 cars. In SAE L3, individuals are more engaged in NDRT and pay less attention to surroundings (Carsten et al., 2012; Naujoks et al., 2016) compared to L2, which may explain why most drivers missed the failure in the present study. Another reason could be attributed to automation complacency, defined as the moment when “a system malfunction, anomalous condition, or outright failure is missed” (Parasuraman & Manzey, 2010). This is likely to happen when individuals are either engaged in a visually demanding task (e.g., word search, surfing the internet, watching videos) or monitoring the automated transport operations (Bailey & Scerbo, 2007) as it impairs their ability to detect anomalies, failures or unanticipated events.

4.2. Resuming control

Only a small number of participants resumed manual control (6 out of 36) after the silent failure, perhaps because only a few noticed it (see Payre et al., 2022). This is compared to 9 people who resumed manual control in the explicit failure, with 44 % of these (4 out of the 9) spending longer than 12 s looking at the ransomware. This result could suggest non-compliance with the NHTSA driver distraction criterion (2013), as the minimum extrapolated success rate should be 87.5 % (21 out of 24 participants). Furthermore, approximately 22 % (2 out of 9) of those who resumed manual control after the explicit failure looked at the HMI longer than 2 s in any one single gaze. Again, this result may not comply with the NHTSA minimum acceptance success rate.

These observations showed that the ransomware attack simulated in this experiment was distinctive as it drew drivers’ attention away from the road to monitor the in-vehicle display. Rather than focusing on the driving task, some of them looked at the in-vehicle screen for more than 2 s in any one single gaze or spend more than 12 s in cumulated fixations. One reason for this could be that drivers took time to understand what was going on and what the message conveyed by the HMI meant: was it a threat, a bogus message or spam? In addition, it probably took drivers a moment to evaluate whether the automated driving system continued to operate successfully, or whether it was compromised by the ransomware. This is congruent with qualitative responses previously collected from the same experiment (Payre et al., 2022). Drivers did not know whether the connected and automated systems in the vehicle were mutually exclusive or intertwined. The distractive effect of the ransomware on drivers may stem from internet users being sometimes over-optimistic concerning their online security (De Kimpe et al., 2021).

While driving, participants seemed to use different strategies to look at the ransomware: few long gazes, and/or many brief gazes sometimes cumulating to more than 12 s spent eyes of road. Either way, the ransomware was distinctive (i.e. drivers look at it instead of focusing on the road) and one driver crashed within seconds after resuming control. Hypothesis 2, supposing that the number and duration of gazes on the HMI are lower in manual than automated driving, is therefore rejected.

4.3. Early vs Late failure

With respect to the timing of failure (i.e., early vs late), we expected that drivers would look more often and longer at the HMI after the early failures as the system appeared vulnerable and compromised sooner rather than later, casting doubt on its reliability (Lee & See, 2004). Based on eye glance behaviours alone, hypothesis 3 was therefore rejected as the number and duration of fixations were similar between timings. One explanation is that most gazes after the failures were directed towards the HMI within a relatively short time window, which is congruent with the gaze patterns measured after each failure and event. The 6 min between the early and late failures may have been too small of a difference to affect drivers’ gaze behaviour, and should be extended in future studies.
4.4. Limitations, implications and future research

One limitation of this experiment is that some drivers might not have thought that the ransomware was genuine, possibly because the comprehension of the ransomware was not supported by the ambiguous nature of the message (Payre et al., 2019). Some of them might have kept in mind they were in a driving simulator, while some others had mentioned they thought it was a spam and did not pay much attention to it. Recent work suggesting that drivers’ perception of risk in automated driving was diminished in simulators supports that assumption (Perelló-March et al., 2022). Despite the limitations of it being a simulator research, the results still showed that the experiment was engaging enough to draw their attention and not ignore the HMI.

The findings have implications for the industry regarding driver distraction and cybersecurity. Firstly, missing the turn signals in a potentially hazardous situation is dangerous and may lead to crashes (Nguyen-Phuoc et al., 2020). In the present experiment, this is not the responsibility of the driver, it is a failure coming from the vehicle, and few drivers noticed this anomaly. Secondly, looking at and interacting with the ransomware while driving was a secondary task. Results showed that two of NHTSA’s distraction criteria to perform that secondary task while driving were not met: total time spent to complete the task greater than 12 s, and single glances longer than 2 s. This means that unsecure automotive HMIs are potentially distractive and can undermine road safety. This is also illustrated in this research where one participant crashed the car shortly after resuming control when the ransomware popped on the HMI. Unsecure HMIs that do not comply with the NHTSA may not be available on the market. However, it is hard to imagine that all HMIs, along with all connected and automated vehicles will be 100 % secure and cyber-attack proof.

While previous traffic simulation work investigated the effect of slight cyberattacks on longitudinal control safety of a CAVs fleet (Li et al., 2018), this study shed light on drivers’ gaze behaviour and distraction. Future research should examine the effect of cyberattacks on take-over performance. The cyber threat in this study was personal data encryption, but what would happen if a cyber-attack threatened to lock the automated driving mode or take control of the vehicle? With respect to the external validity, the present study did not investigate how drivers would react in more complex and demanding environments, for instance when interacting with different road users (e.g., cyclists, pedestrians) in busy built-up areas. Finally, more ecological settings such as on-road trials will further our understanding of drivers’ responses to silent and explicit failures.

5. Conclusions

The answer to the research question “Are silent and explicit in-vehicle screen failures in CAVs an emerging source of distraction for drivers?” is yes with respect to explicit failures resulting in a ransomware popping on the HMI. The ransomware was distracting because it led some drivers to shift their attention away from the road. Echoing the ironies of automation (Bainbridge, 1983), this simulator research highlights one paradox stemming from CAV: these vehicles are expected to be safer and more efficient than human drivers, but ultimately, they may not be the panacea for road safety ills as they fail to mitigate one of the main causes of road accident, namely distraction.

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.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the UKRI Trustworthy Autonomous Systems Hub (grant number: EP/V00784X/1). The authors would like to thank Hesamaldin JadidBonab whose technical expertise help develop the real-time link between the simulator and the HMI, and also the cyber-security co-investigators on the project: Giedre Sabaliauskaite and Siraj Shaikh.

References


