ORIGINAL ARTICLE

OESDs in an on-road study of semi-automated vehicle to human driver handovers

Neville A. Stanton[1] · James W. Brown[1] · Kirsten M. A. Revell[1] · Jisun Kim[1] · Joy Richardson[1] · Pat Langdon[2] · Mike Bradley[3] · Nermin Caber[3] · Lee Skrypchuk[4] · Simon Thompson[4]

Abstract

Design of appropriate interaction and human–machine interfaces for the handover of control between vehicle automation and human driver is critical to the success of automated vehicles. Problems in this interfacing between the vehicle and driver have led, in some cases, to collisions and fatalities. In this project, Operator Event Sequence Diagrams (OESDs) were used to design the handover activities to and from vehicle automation. Previous work undertaken in driving simulators has shown that the OESDs can be used to anticipate the likely activities of drivers during the handover of vehicle control. Three such studies showed that there was a strong correlation between the activities drivers represented in OESDs and those observed from videos of drivers in the handover process, in driving simulators. For the current study, OESDs were constructed during the design of the interaction and interfaces for the handover of control to and from vehicle automation. Videos of drivers during the handover were taken on motorways in the UK and compared with the predictions from the OESDs. As before, there were strong correlations between those activities anticipated in the OESDs and those observed during the handover of vehicle control from automation to the human driver. This means that OESDs can be used with some confidence as part of the vehicle automation design process, although validity generalisation remains an important goal for future research.

Keywords OESD · Handover · Driving automation · Interface design · Validation

1 Introduction

The design of the interface between human driver and vehicle automation will be critical for successful handovers, especially when control of the vehicle is being handed back to the human driver (Eriksson et al. 2019; Eriksson and Stanton 2017a, b; Clark et al. 2019a). There are many examples of the failure to hand vehicle control back successfully, from both simulator studies (Banks et al. 2014; Stanton et al. 1997) and real on-road collisions that have ended in fatalities (Banks et al. 2018a; Stanton et al. 2019). The problem of handing back control from an automated system to a human operator is not restricted to the road domain, as examples from aviation have shown (Stanton and Marsden 1996; Salmon et al. 2016). In aviation, mode confusion in a perfectly functioning aircraft has led to situations whereby pilots have crashed. Air France flight AF447 is a case in point. The hand-back of control of the aircraft was misunderstood by the pilots, who seemed to think that the aircraft was in an over-speed condition. This is a classic example of mode confusion (Sarter and Woods 1995). The confusion led the pilot flying to pull back on the control stick, which induced a wing stall in the aircraft, causing the aircraft to drop into the ocean killing all on board (Salmon et al. 2016). Mode confusion is also of great concern in automated road vehicles (Stanton et al. 2011). When vehicle control is handed back to a human driver, it is particularly important that the driver is aware of the vehicle status, road environment and pertinent road infrastructure as well as other road users (Stanton et al. 2017). Therefore, the design of the

✉ Neville A. Stanton
   N.Stanton@soton.ac.uk

1 Human Factors Engineering, Transportation Research Group, Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton SO16 7QF, UK

2 Edinburgh Napier University, Edinburgh, UK

3 University of Cambridge, Cambridge, UK

4 JaguarLandRover, Warwick, UK

handover is critical for success (Banks et al. 2018b; Clark et al. 2019b; Eriksson et al. 2019).

A structured approach to design is required for developing the handover requirements and Operator Event Sequence Diagrams (OESDs) are one such approach amongst others (Stanton et al. 2013). OESDs were selected for this project because they had been used previously in simulator studies of driver handovers (Stanton et al. 2021). OESDs have been used in the design of human–machine interaction and interfaces for over 60 years (Kurke 1961). Since that time they have been used in a wide variety of applications, including analysis of aircraft landing procedures (Sorensen et al. 2011), evaluation of single pilot operations (Harris et al. 2015), analysis of the relationship between air traffic control and flight decks on civilian airliners (Walker et al. 2010), analysis of the activities between maintenance operator in the field and operations in the central control room of an electrical energy distribution company (Salmon et al. 2008), comparison of traditional and new procedures for managing collision avoidance in the maritime domain (Kurke 1961), and comparing approaches to automatic emergency braking systems in road vehicles (Banks el al. 2014). OESDs represent the different aspects of systems (including the human operator, interfaces and technical elements) in separate columns (colloquially called 'swim-lanes' in OESD parlance), against time.
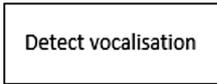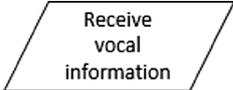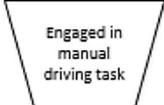
Previous studies into the design of handover protocols for vehicle automation have found that the OESDs have good predictive validity, with median Phi of over 0.8 (Stanton et al. 2021). This means that the OESDs were able to predict the majority of the behaviours drivers engaged in during the vehicle control handover process. In both of these studies, OESDs were constructed in workshops with experts in vehicle engineering, computing and human factors. The behaviour of the driver during that handover process was described in the driver column together with any interactions with other aspects of the vehicle (such as hearing or reading any information as well as making any verbal or physical responses). The drivers' behaviours during the hand-back of vehicle control in the OESD were then compared to those observed in the video recordings from the driving simulators. In the first study there were over 100 drivers in two separate driving simulators, one desktop ($N=49$) and one full vehicle ($N=60$). In both studies, the predictive validity of the OESDs was very good (Stanton et al. 2021). In the second study, there were 65 drivers undertaking 4 trials and the correlational data showed there was also a good predictive validity for each trial (Stanton et al. 2021). As both of these previous studies report on data collected in driver simulators, so it is an important to see if these findings generalise to on-road studies. Whilst it may seem reasonable to expect the predictive validity of OESDs for the hand-back of vehicle control to human drivers in on-road studies to mirror

those of driving simulators, it is important to test validity rather than assume it (Stanton and Young, 1999; Stanton 2016). To that end, the OESDs were constructed prior to the on-road studies, as described in the next section.

2 Construction of OESDs

Further guidance on the development of OESDs may be found in Kirwan and Ainsworth (1992) and Stanton et al (2013). The analysis presented in this paper is based on a use-case of vehicle automation handover scenario on a UK motorway with a SAE level 3 vehicle (SAE 2016). It is assumed that drivers will drive manually onto the motorway and hand the driving task over to vehicle automation when it becomes available. Whilst vehicle automation is engaged, the driver is free to engage in non-driving tasks (such as reading, emailing, working on a tablet computer). The vehicle would alert the driver of the need to take back control of the vehicle in a planned, non-emergency, handover in a timely manner before the exit junction. These handovers are described using the task elements from Operator Event Sequence Diagrams as shown in Table 1.

Table 1 Key for the Operator Event Sequence Diagrams

OESD task elements	Description
Detect vocalisation	A generic task
Receive vocal information	Information is received
Intelligible Response? (No / Yes)	A decision with two or more possibilities
"Hazard vocalisation"	Audio output
Autopilot icons unlit	Display output
Engaged in manual driving task	A manual task

The Operator Event Sequence Diagram shown in Figs. 1, 2, 3, 4, and 5 were developed in workshops with Human Factors and Automotive Engineering experts. The 9 swimlanes show in Fig. 1 the different 'actors' under consideration in the design of the handovers to and from the human driver and vehicle automation (via the instrument cluster—instruments viewed through the steering wheel), HUD (Head-Up Display viewed in the windscreen or windshield), centre console (the upper part of the centre of the dashboard), ambient (lighting around the dashboard and vehicle interior) and haptic (vibration through the driver's seat) displays). The arrows are connectors that show the links between the events in the swimlanes. The handover protocol presented in Figs. 1, 2, 3, 4, and 5 were designed to raise the situation awareness of drivers, by presenting them with contextually relevant questions about the vehicle status, other road users as well as the surrounding environment and infrastructure. This was based on the research evidence that degraded performance of drivers of automated vehicles is, in part, due to

poor situation awareness (Stanton et al. 2017). For example, the collisions in the Tesla and Uber vehicles report that the driver was not aware of the environment outside the vehicle (Banks et al. 2018a, b; Stanton et al. 2019). As can be seen in Fig. 1, it is assumed that the vehicle is under manual control until the system detects that the road is suitable for automation to operate. Then the system prompts the driver with the message that automation is available, via the four interfaces (cluster, HUD, centre console and ambient display) should they wish to use it.

If the driver chooses to engage vehicle automation, then they would press two buttons on the steering wheel with their thumbs simultaneously (assuming that their hands are in the ten-to-two clock position on the steering wheel). At this point, the interfaces would display "Automation Activated" followed by "The car is in control" (see Fig. 2). At the same time, the ambient lighting in the car would change from orange (indicating manual driving mode) to blue (indicating automated driving mode). Then, the driver is able to
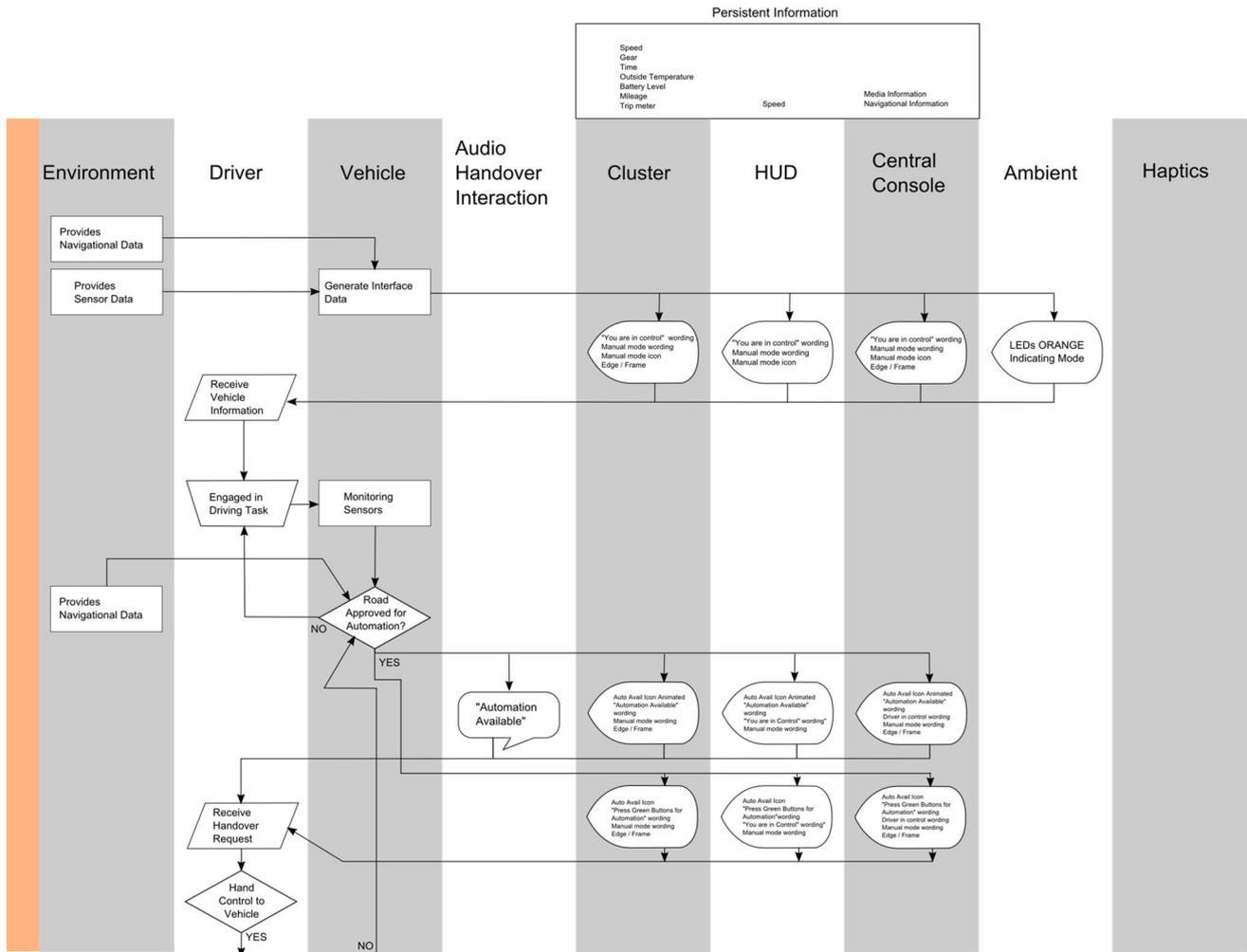

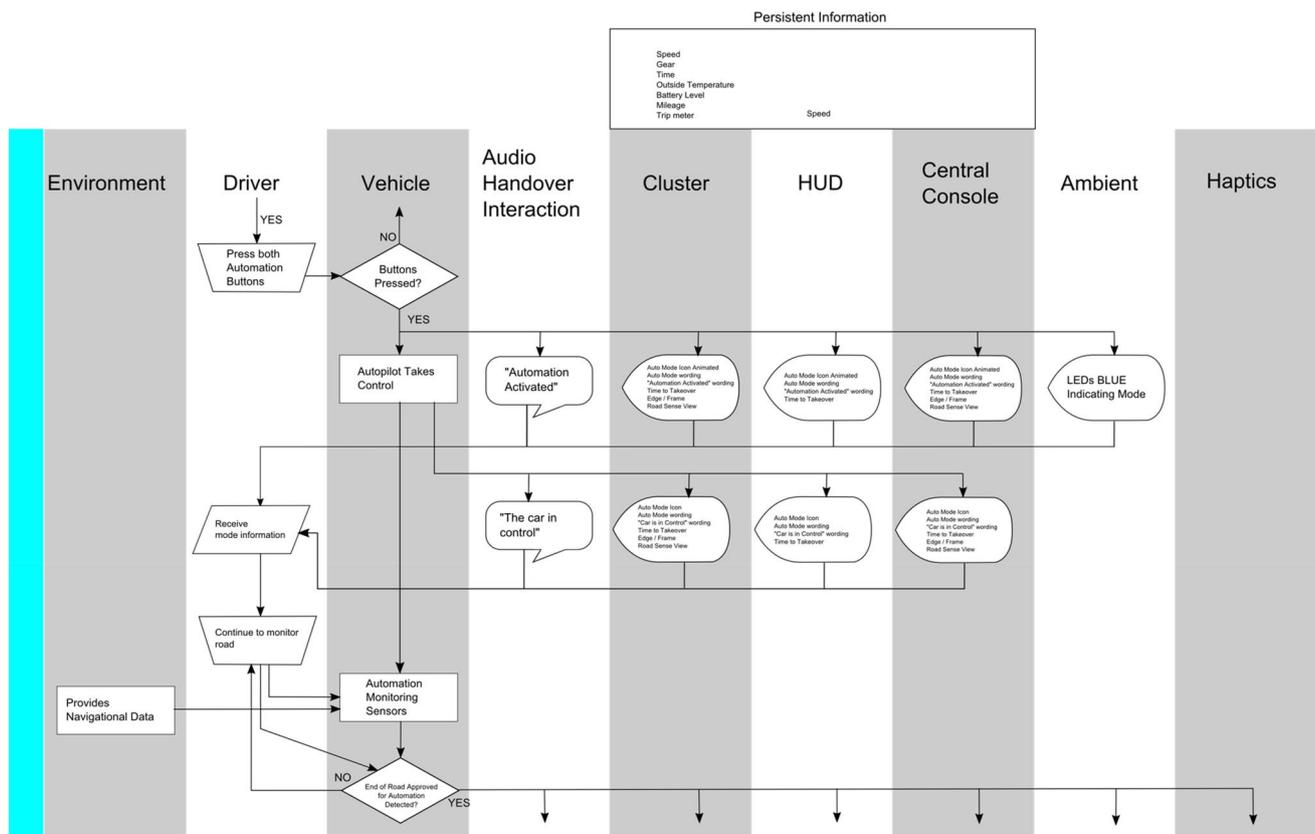**Fig. 1** Vehicle in manual mode with automation available

Fig. 2 Vehicle transferred to automated mode by human driver

engage in non-driving tasks (on a tablet computer in this scenario, as SAE (2016) level 3 is assumed, so there is no need for the driver to monitor the automated driving system). The scenario assumes that there is a planned handover of driving from automation back to the human driver (such as when their exit from the motorway is coming up, which would have been pre-programmed into the satellite navigation system). The driver is given a 5, then 2, then 1 min notice that the handover process will begin (see Fig. 2).

Upon the prompt from the automated system that the driver needs to get ready to drive, it is assumed that the driver ceases the non-driving task, puts down the tablet computer, and resumes the driving position (as shown in Fig. 3). The system then presents a series of questions designed to raise the situation awareness of the driver (such as: what speed is the vehicle currently travelling at? What lane are you currently in? What colour is the vehicle in front of you? What is your remaining fuel range? Can you see a bend in the road ahead?). The driver is expected to respond to these questions (which are presented auditorily as well as on all of the visual interfaces). If the answer is correct, then the next question is presented until all questions have been answered. If the answer is incorrect, then the question is repeated for a maximum of two additional times before moving onto the

next question. When all questions have been presented, the handover interaction moves on to that presented in Fig. 3.

The driver is then requested to take manual control of the vehicle, which will mean placing both hands on the steering wheel and positioning their foot on the accelerator pedal. To transfer control from the automated system to the driver, they need to press two buttons mounted on the steering wheel at the ten-to-two clock position with their thumbs (in the same manner as they do for handing control over to the vehicle automation system), as shown in Fig. 4.

As Fig. 4 shows, the driver is then requested to take manual control of the vehicle, which will mean placing both hands on the steering wheel and positioning their foot on the accelerator pedal. To transfer control from the automated system to the driver, they need to press two buttons mounted on the steering wheel at the ten-to-two clock position with their thumbs (in the same manner as they do for handing control over to the vehicle automation system).

When control of the vehicle is passed back to the human driver, the ambient lighting changes back from blue to orange (indicating the vehicle is now in manual driving mode) and the words "Automation deactivated" are presented auditorily as well as on the visual displays. This is followed by the words "You are in control", which are also
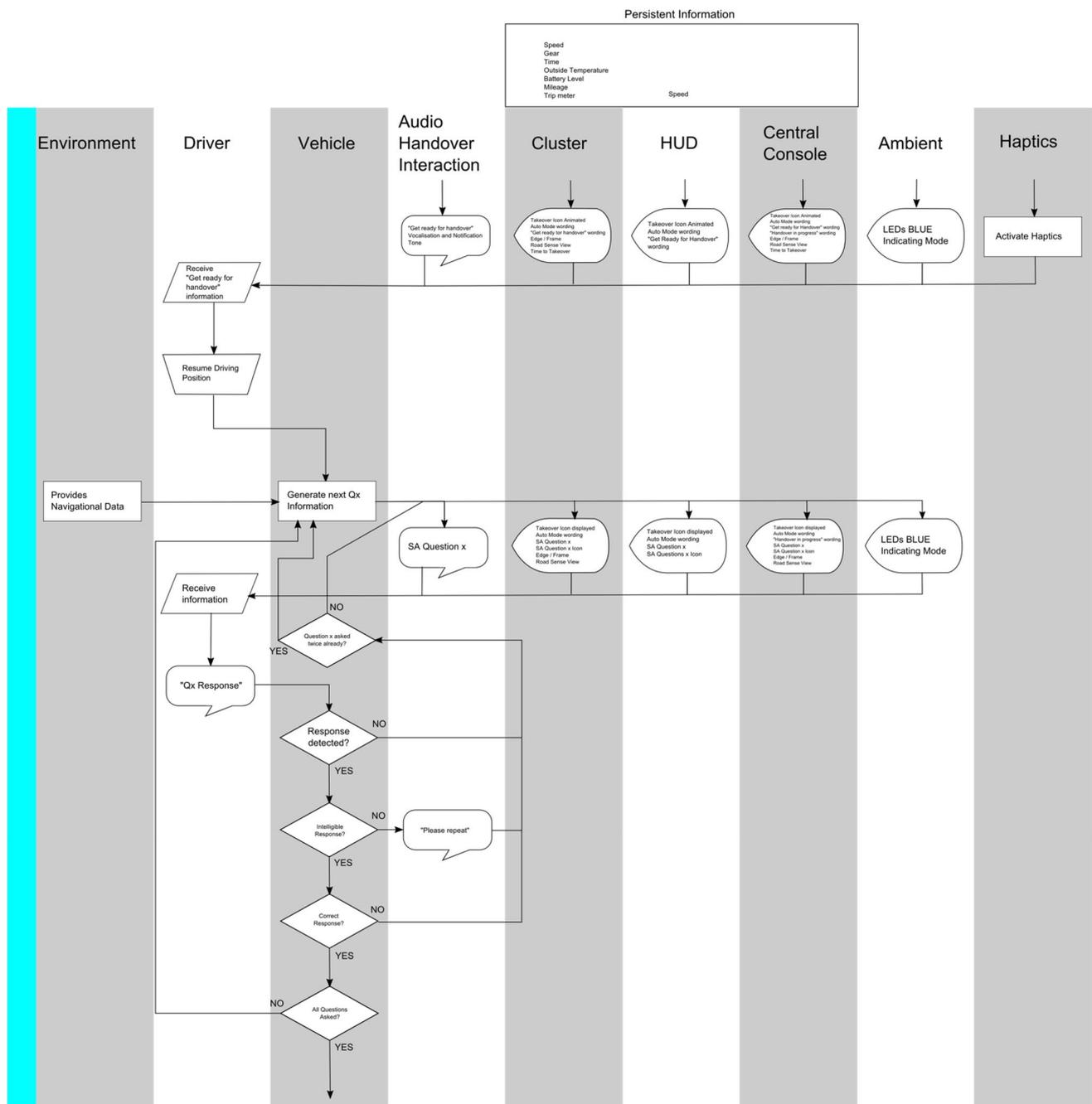
Fig. 3 Vehicle in automated mode preparing to transfer to manual mode

presented auditorily and on the visual displays (as shown in Fig. 5). The human driver is now driving the vehicle.

The next section describes the study in which the video data was collected from human drivers of an automated vehicle on the road. These data were used to validate the handover from vehicle automation to human driver, as show in Figs. 3, 4, and 5. The scientific contribution of this paper is the validation of OESDs in the context of the design of handover interaction and interface design. The

research also aimed to demonstrate how the design of the handover between humans and machines could be facilitated by the use of OESDs. In addition, on-road studies of the nature undertaken in this research are exceedingly rare at the moment.
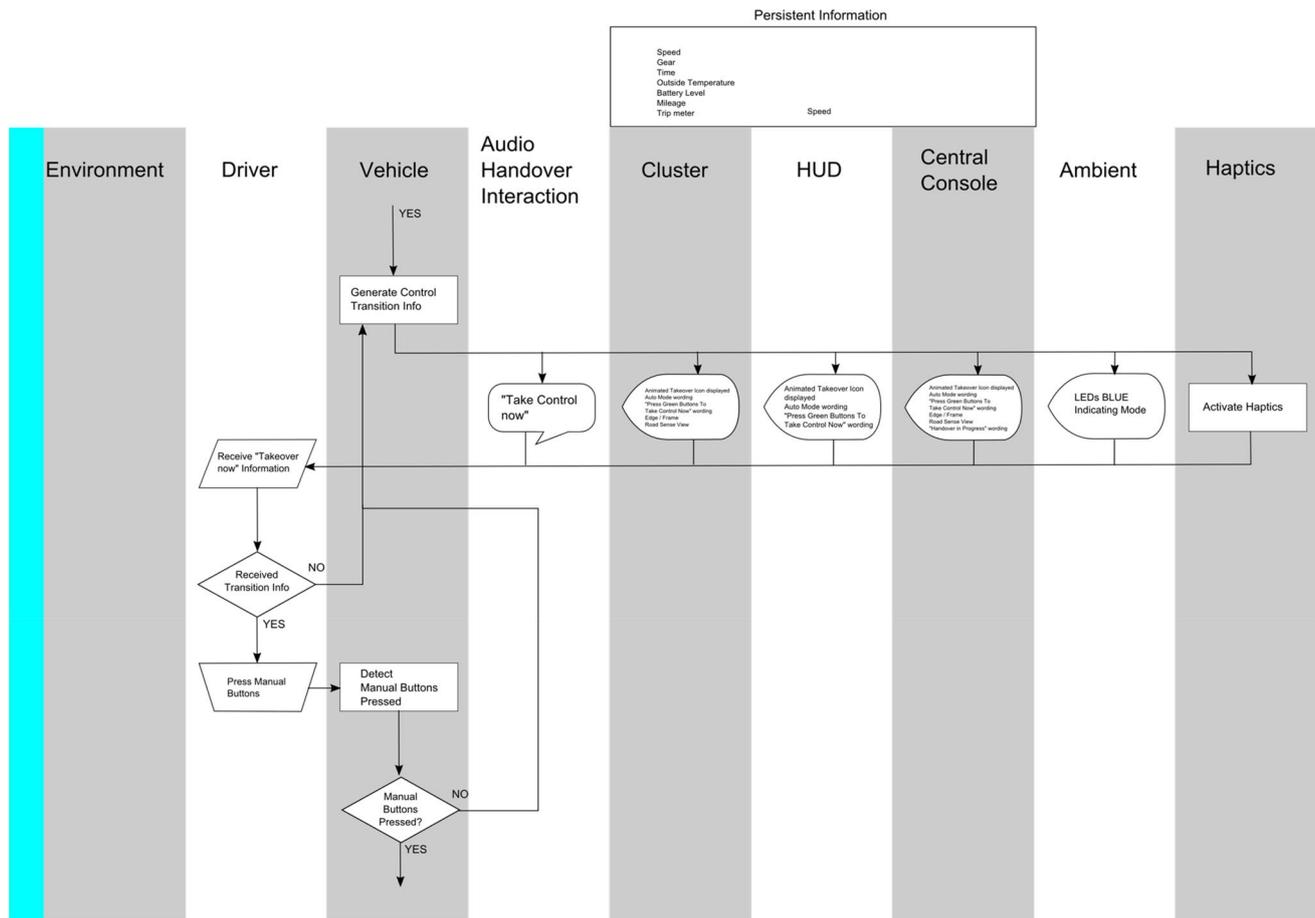
Fig. 4 Vehicle transferring from automated mode to manual mode

3 Methods

The experimental methods used in this study covers the participants, experimental design, equipment, procedure, data reduction and analysis.

3.1 Participants

Although 24 participants were recruited from a pool that had previously taken part in a simulator-based study featuring a similar HMI, data from only 16 was good enough to be used in this study due equipment and recording failures. All drivers held full UK driving licences. There were 10 males and 6 females. The age range of drivers was 29–67 years (mean = 46.3 years, SD = 11.7 years). Participants provided signed consent prior to involvement in the study that was approved by the University of Southampton's Ethics and Research Governance Office (ERGO 49792.A2).

3.2 Experiment design

A repeated measures design was employed for the experiment, covering three handover events using the default HMI settings, as shown in Fig. 6.

The Driving Mode Wording is situated below the Driving Mode Icon, this is centrally placed in the infotainment display and cluster, and placed on the left side in the HUD. The colour acts as a mode indicator, blue indicating automated and orange for manual. Time to Takeover is shown in the bottom left corner of the Cluster and Infotainment Display, and the top right of the HUD. The Edge Frame mode indicator is only visible on the Cluster and Infotainment Display. Handover Questions and associated icons appear in the centre of each of the HMI display elements (Fig. 7).

Video cameras were positioned around the interior cabin of the vehicle to record the driver interaction with vehicle automation so that it could be compared with the OESDs.
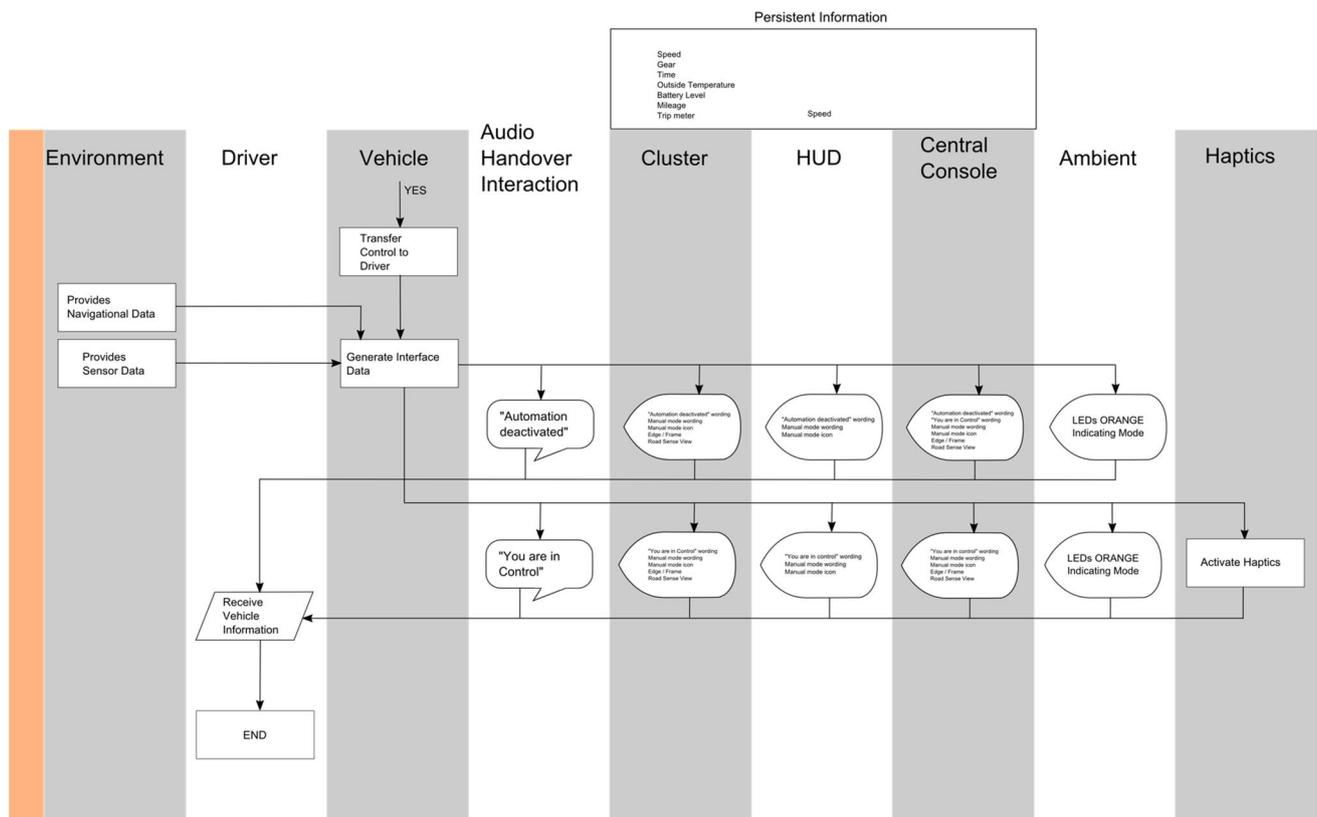
Fig. 5 Vehicle back in manual mode

Fig. 6 The customisation matrix displayed at the end of each trial on the vehicle's infotainment display

3.3 Equipment

The experimental vehicle was a 2017 Jaguar iPace EV400 AWD pre-production model. The automation system consisted of a combination of factory standard Lane Keep Assist (LKA) and Adaptive Cruise Control (ACC). When utilising these systems, the standard car would issue a frequent warnings to maintain hands on the wheel, these warnings were specifically removed from the system for it to simulate SAE level 3 automation. The visual aspects of the HMI consisted of a 14″ TFT panel fitted to the centre console, a 10″ TFT fitted in place of the OEM cluster, and a HUD comprising a small 1000 × 250 px TFT and reflector screen (Bysameyee 2020). The car's
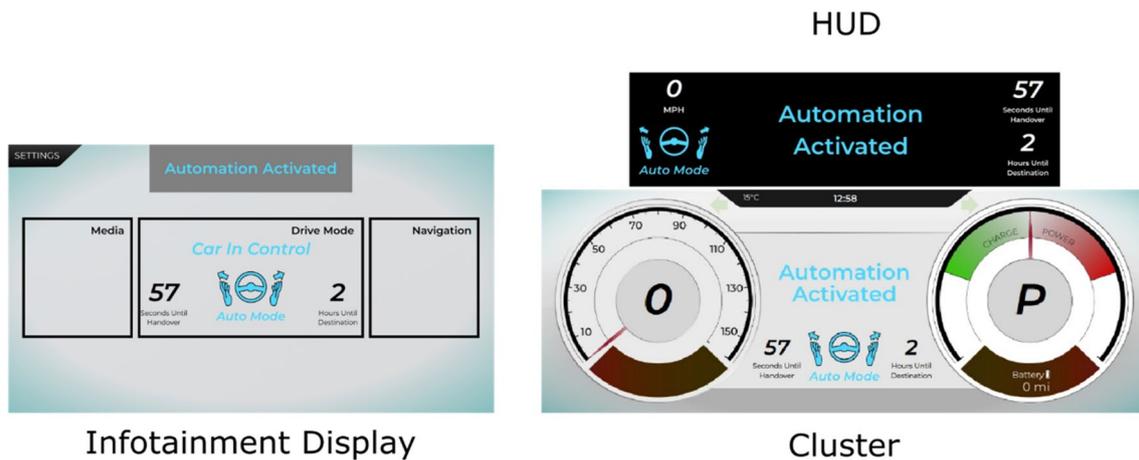
Fig. 7 This interfaces in automated mode, showing centre infotainment display (left), head-up display (top right) and instrument cluster (bottom right)

OEM cluster was moved into the passenger footwell to allow monitoring by the safety driver. The interactive element of the HMI constituted two illuminating green buttons fitted in the thumb positions of the steering wheel (approximately at the 10–2 clock position). Ambient lighting was supplied via LED lighting strips. Haptics were provided in the seat base via Leggett & Platt motors controlled through an Arduino Micro and motor control board. Five KT&C cameras were installed within the iPace to provide footage from the forward view, driver facing, over-the-driver-shoulder, footwell and rear views. A dashcam was also fitted in the safety car to collect footage of the participants' vehicle from an external viewpoint. The test vehicle was a 2017 Jaguar

iPace EV400 AWD and Fig. 8 shows the car in manual mode with the HUD and cluster visible.

In addition, the n-back task presented a continuous stream of digits aurally to the driver who was required to repeat 1-back from the digit being presented (AgeLab 2019). This was used a way of loading the working memory of the driver to cognitively distract from the driving task. Ideally, when in automated mode, the driver would have been distracted visually from the road environment but this was not allowed due to the risk assessment. This was the reason for introducing the n-back task. It was reasoned that this was a cognitive distraction (albeit not a visual one). Nevertheless, it did enable the study to go ahead as the project faced significant technical. legal and ethical challenges to undertake a study
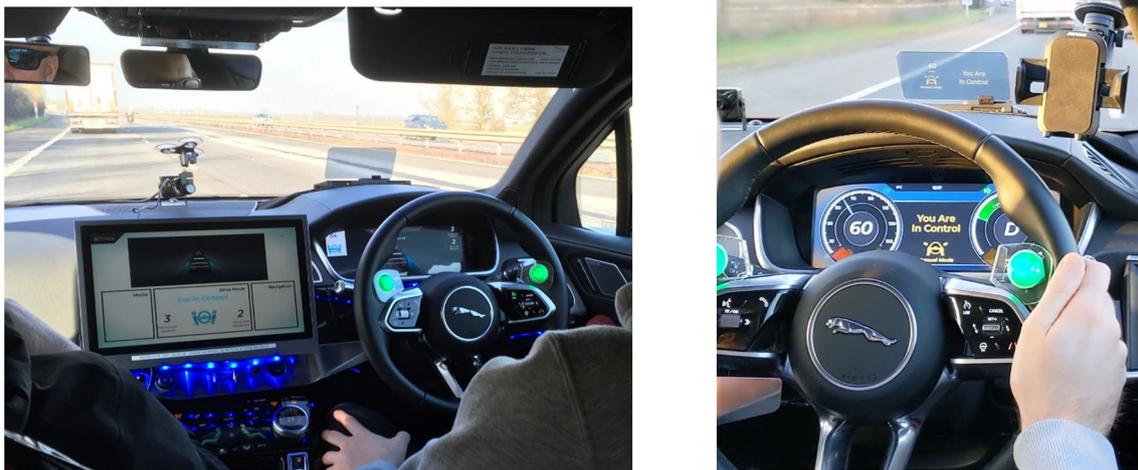

Fig. 8 The left-hand image shows the green buttons on the steering wheel and the infotainment display (ambient blue lighting indicates that the vehicle is in automated mode). The driver-facing camera can be seen fitted on the dashboard. The figure on the right shows the car in manual mode with the HUD and instrument cluster

on British motorways. In any case, we were studying the planned hand-back of the vehicle control from automation to the human driver, in which it was anticipated would involve a period of the driver sampling the road environment before the handover began AgeLab (2019).

3.4 Procedure

On arrival at the JLR facility at Fen End, the participant was welcomed and asked to sign in, presenting their driving licence and receiving a visitor pass. They were then briefed on the aim of the project and how the on-road study follows on from the simulator study that they completed the previous year. A brief explanation of the sequence of events for the study was then given. This included the route, the takeover procedure, form filling on trial completion, customisation of the HMI and the n-back cognitive load task (to simulate the driver undertaking another (verbal) task whilst the vehicle was being driven by the automation). It was stressed that they were required to maintain the same level of attention as if they were driving and be ready to take control at all times. They were also advised that if the cognitive load task was detrimentally affecting their ability to retain a safe level of attention, they should stop the n-back task. They were presented with a copy of the participant information sheet, this same sheet was made available to them when they were recruited. A reminder sheet showing screenshots of the HMI they would be using was provided, including modal variation of the cluster, HUD and IVI, the green buttons and the customisation matrix screen. They were shown a route map that highlighted where the trials would take place, but also areas of caution and one junction that required traversal in manual mode. A privacy policy was presented, and a sheet explaining the cognitive load n-back task was provided along with a brief verbal explanation. The participant was asked if they had any questions, and if they were happy to continue, they were provided with two consent forms, an attendance form and an events team form to sign. On completion of all forms, they were lead from the reception area to the car park, and asked to sit in the passenger side of the car, both experimenters took their places in the rear seats. The participant was introduced to a safety driver who explained the basic controls on the car, and the elements of the interface. The safety driver then drove the car through security to the proving ground and demonstrated the vehicle's performance before running through the transition to automated mode and back to manual control. The automation system was operated by one of the experimenters using a Wizard of Oz system, from a Windows tablet in the back of the car. To offer automation, an experimenter would press a start button on a custom control panel app running on the tablet; the HMI then indicated automation as being available via the three graphical interfaces and a vocal alert. To enable automation, the driver simultaneously pressed the two green buttons mounted on the steering wheel and released all of the controls, including the accelerator. The system would then engage automation and the HMI would indicate that the automated mode was active. During automation, the cognitive load task was controlled by the other experimenter via a mobile telephone-based app linked to a Bluetooth speaker. The safety driver demonstrated the automation system multiple times including the n-back task, and requested the safety car overtake and brake in front of the vehicle, when in automated mode, to illustrate how it reacts to maintain a gap to the car in front. The safety driver then stopped in a safe area and swapped places with the participant. The participant then was allowed some time to drive the car on the proving ground to become familiar with the controls. The automation was then made available to them and they experienced multiple handovers and takeovers, including the use of the n-back task whilst in automated mode.

The participant was asked to drive manually to the start point of the experiment at the Southbound Warwick services on the M40. Two miles prior to the services, whilst on the M40, a road-based practice handover was conducted. The automation was offered to the participant, once activated, a short 30 s period of automation followed which included the participant carrying out the n-back task. The participant then experienced the handover protocol and resumed manual control before stopping at the services. After confirming that the participant was happy to continue, the on-board systems were checked and configured for the first trial. data logging was started, and the video and audio was synchronised using a clapper board.

The participant was instructed to drive from the services onto the M40, proceeding in the left lane at approximately 58mph. After 1 min of manual driving, automation was offered to the participant. Once activated, after a further period of 30 s, the n-back cognitive loading task was started. This task was run in 2 min intervals, separated by 30 s. Following 10 min of automation, the HMI started the takeover protocol, once completed, the participant pressed both steering wheel buttons and resumed manual control. After one minute of manual control, automation was again offered to the participant and the process repeated. On completion of the second handover, the manual driving period was extended to approximately 7 min, this was to pass a section of motorway (M40 J9 Southbound) that would have adversely affected the automation due to the lane becoming a sliproad. Due to proximity to the end point of trial 3, the automation period was reduced from 10 min to 8 min. The n-back task was started simultaneously with the automation for 30 s, before reverting to 30 s off and 2 min on until the protocol started. When the participant completed the protocol and assumed manual control, they continued in manual mode the short distance remaining to the motorway services

and parked the car. This concluded the data collection the trial and safety driver swapped places with the participant from this point and drove the vehicle back to the JLR facility at Fen End. The participant was thanked for their time and signed out at reception.

3.5 Data reduction and analysis

The validation of the OESDs was assessed by comparing the video data collected on the road of the driver during the automation handover process with those tasks identified in the OESDs. In the OESDs, 15 tasks were defined as follows:

1. Receive "Get Ready to Take Over" Information
2. Resume Driving Position
3. Receive SA Question 1
4. Answer Question 1
5. Receive SA Question 2
6. Answer Question 2
7. Receive SA Question 3
8. Answer Question 3
9. Receive SA Question 4
10. Answer Question 4
11. Receive SA Question 5
12. Answer Question 5
13. Receive "Take Control Now" Information
14. Received Transition Information
15. Press Manual Buttons

These data were processed using the signal detection paradigm (Green and Swets, 1966) which discerns between four events: Hits, Misses, False Alarms (FA), and Correct Rejections (CR). In the context of this experiment, it provided a method by which to compare predicted driver behaviour illustrated on OESDs with driver behaviour, observed during the on-road trials of the vehicle control handover process (as shown in Fig. 9).

The four events in Fig. 9 are defined for all driving-related tasks as follows:

Hits: present in Operator Event Sequence Diagram and present in the video of automation-driver handover.

Misses: not present in the Operator Event Sequence Diagram but present in the video.

False alarms: present in the Operator Event Sequence Diagram but not present in video.

Correct rejections: not present in Operator Event Sequence Diagram nor in the video (Unique misses by all participants minus the individual participant's misses).

The latter category can be difficult to calculate as it could be infinity, but for the purposes of this investigation, it should be based on the total number of unique misses generated by all of participants, minus the number of misses

Fig. 9 Signal detection theory (SDT) matrix

for each individual participant. Additional the Hit Rate and False Alarm Rate were calculated as follows:

Hit Rate = Hits/Hits + Misses,

False Alarm Rate = False Alarms/False Alarms
+ Correction Rejections.

Inter-rater reliability testing was conducted on the categorisation scheme for approximately 20% of the video footage between two analysts. Equal weighted Cohen's Kappa was calculated (0.781) showing acceptable agreement between the two independent analysis in their classification of hits, misses, false alarms and correct rejections (Landis and Koch 1977).

Finally, Matthews (1975) correlation coefficient (Phi—a correlation coefficient for dichotomous data) was applied to the data generated by the SDT analysis, this quantified the correlation between the expected and observed behaviour, as a means to validate the OESDs. The Matthews correlation coefficient formula is shown as follows:

$$\varphi = \frac{\text{Hit} \times \text{CR} - \text{FA} \times \text{Miss}}{\sqrt{(\text{Hit} + \text{FA})(\text{Hit} + \text{Miss})(\text{CR} + \text{FA})(\text{CR} + \text{Miss})}}.$$

Phi was selected as an appropriate statistic for validation of Human Factors methods, as it has been used previously with some success (Stanton and Young 1999, 2003). In particular, the method of categorising the data offers a way of correlating predicted driver behaviour with that observed in the videos.

4 Results

The data from the three handovers from the automated vehicle to the human driver are summarised in Table 2 for each participant. As there were 15 tasks per handover, 45 hits represent a perfect score, meaning all of the tasks in the OESD were observed in the video. One participant had 13 False Alarms because they cut the handover short by assuming control of the vehicle without passing through all of the preceding tasks (participant 22). The misses are shown in Table 3 with the frequencies by handover. As there was a total of 6 unique misses per handover (observed from the video data), the total correct rejections were 18 if none were observed. Hit rate, false alarm rate and Phi were calculated as described in the data reduction section.

The misses are an interesting category of events, as they describe activities that were not predicted by the Operator Event Sequence Diagram but were present in the video of the driver. Over half of the misses (15/29) related to the driver covering the green buttons on the steering

wheel with their thumbs (the 10–2 position), early on in the handover process. This is not a safety concern, however, as shows that the drivers were readying themselves to resume manual control. On one occasion, the driver failed to assume the driving position at the right point in the process (task 2), as they remained with their hands on their lap and their feet away from the pedals. The only consequence of this is that it increases vehicle control handover time, but as this time is entirely driver paced there is no adverse consequence for safety. On one occasion, a driver (participant 22) took control over the vehicle early in the process, which shorted the handover protocol considerably. Whilst any handover protocol should not hamper drivers of be overly intrusive, there is a fine balance of ensuring that the drivers awareness has been raised sufficiently for them to resume control of the vehicle (Stanton et al. 2017). Six drivers did not assume driving position until right at the end of the handover protocol, whereas it had been anticipated that they would be in this position from the start. Again, they were no safety implications from this behaviour. Five participants removed their hands

Table 2 Presentation of data by participant (rounded to two decimal places*)

Participant	Hits	Misses	FA	CR	Hit rate*	FA rate*	Phi*
1	44	1	1	17	0.98	0.056	0.92
2	44	8	1	10	0.85	0.09	0.63
3	45	2	0	16	0.96	0	0.92
4	45	3	0	15	0.94	0	0.88
7	45	0	0	18	1	0	1
8	45	1	0	17	0.98	0	0.96
9	45	3	0	15	0.94	0	0.88
12	45	1	0	17	0.98	0	0.96
13	45	0	0	18	1	0	1
17	45	1	0	17	0.98	0	0.96
19	45	3	0	15	0.94	0	0.88
20	45	3	0	15	0.94	0	0.88
21	45	0	0	18	1	0	1
22	32	3	13	15	0.91	0.46	0.49
23	42	0	3	18	1	0.14	0.89
24	45	0	0	18	1	0	1

Table 3 Misses in the three handovers from vehicle automation to human driver

Misses	Handover 1	Handover 2	Handover 3	Total
Failed to assume driving position	1	0	0	1
Covered buttons with thumbs	6	4	5	15
Took control early	1	0	0	1
Assumed driving position at end	1	1	4	6
Removed hands during protocol	1	1	3	5
Assumed driving position early	0	1	0	1
Total	10	7	12	29

from the steering wheel during the course of the handover, having previously placed them on the wheel. Finally, just one driver assumed the driving position before the protocol had started. In summary, none of these misses presented any safety concerns to the study, particularly as the handover was at the paced of the driver rather than the vehicle automation.

As Fig. 10 shows, most of the hits were almost at the maximum level ($n = 45$), except to 2 participants, most notably participant 22, who resumed manual control early from the automaton system for one of the 3 handovers, which also generated the most false alarms. Participant 2 is less of an outlier, being just below the mean for the hits and just above the mean for the false alarms in the box-and-whisker plots. Participant 2 also has more misses, and consequentially fewer correct rejections, than the other participants. Misses were generally quite low and correct rejections were high for most participants.

Given the findings in Table 2, it is no surprise that the hit rate is high and the false alarm rate is low, as shown in Fig. 11. Again, there are the outliers of participants 2 and 22 for false alarm rate, with the latter being more pronounced. This mean that missing task steps out in the return of control to the human driver is more likely to affect the predictive validity of the OESDs.

Finally, Phi (the correlation coefficient for dichotomous data) averaged from all the individual coefficients, as shown in Fig. 12, reveals good predictive validity for OESDs, above 0.8 (which is generally considered to be the criterion for correlational data). The two outliers are again participants 2 and 22. The latter being the lower Phi coefficient.

In summary then, this means that OESD have been shown to be good predictors of the observed handover behaviours when going from automated control to manual control in 14 of the 16 cases.

5 Discussion

The main finding from the on-road study of handovers from vehicle automation to human drivers is that the OESDs predicted most of the observed activities. To that end, the findings of the study mirrored those from simulator studies undertaken previously (i.e. median Phi > 0.8), albeit with a much smaller sample (Stanton et al. 2021). In addition, all handovers from vehicle automation to human driver were successful, in that they largely followed the protocol designed in the OESD, with the exception of two participants (numbers 2 and 22) who, for one of their three handovers, took control of the vehicle back early. Nevertheless,

Fig. 10 Box-and-whisker plots for hits, misses, false alarms (FA) and correct rejections (CR)
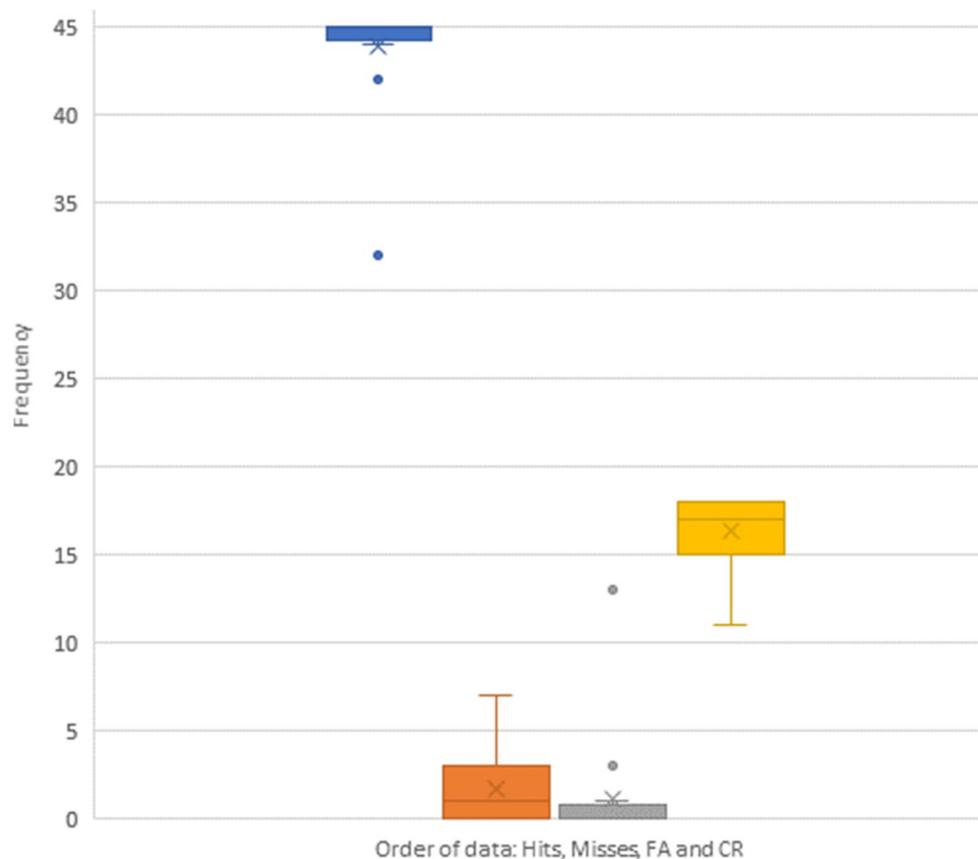
Fig. 11 Box-and-whisker plots for hit rate and false alarm (FA) rate
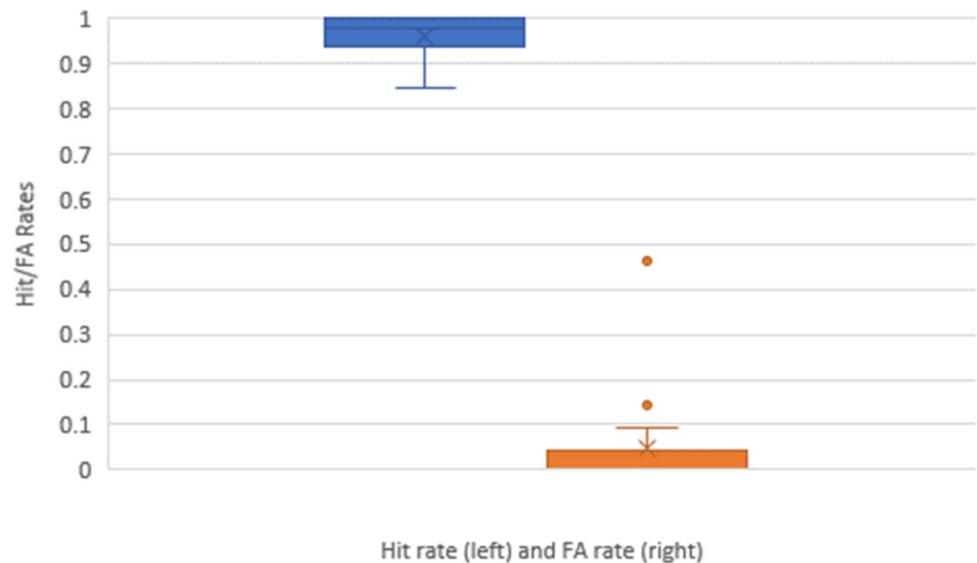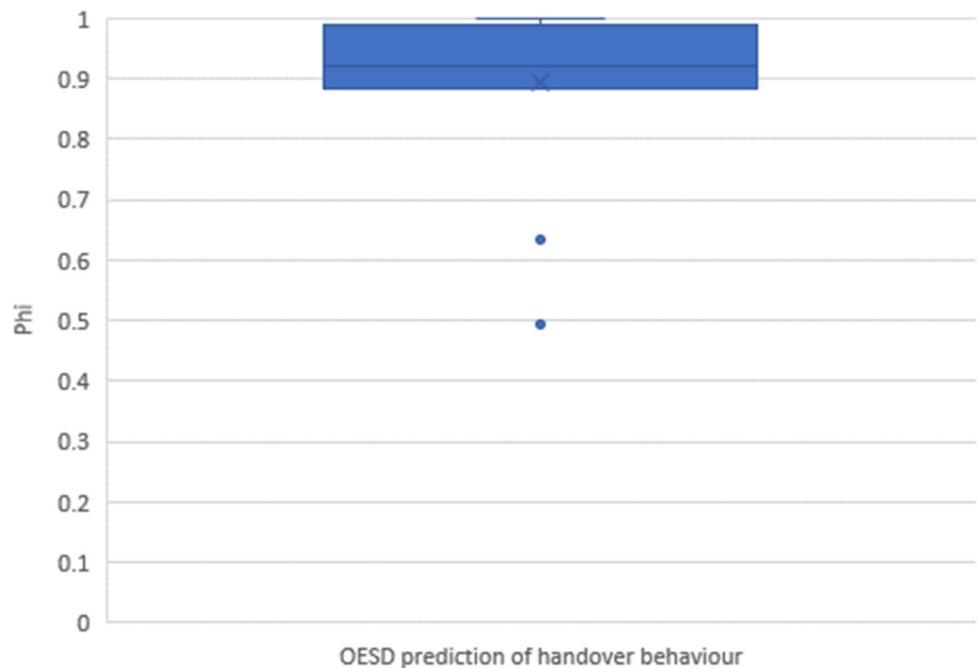

Fig. 12 Box-and-whisker plot for Phi

there was no evidence of mode confusion in any of the handovers (Sarter and Woods 1995; Stanton et al. 2011). Therefore, as an exercise in design, we may conclude that OESDs was a useful Human Factors method (Kurke 1961; Kirwan and Ainsworth 1992; Stanton et al. 2013), and was able to predict automation-driver handover behaviour. There is a contemporary discussion within the Human Factors community about whether or not 'work as imagined' matches 'work as done' (Hollnagel 2015; Read et al. 2018). Therefore, although we used the OESD to design the interaction and interfaces, there is no guarantee that the drivers would behave in a way we imaged they would in the OESD model.

In fact, two of them did not, which proved the point. That 14/16 drivers did indeed behave the way we 'predicted' they would (in the sense that the OESD is both a model of the system and a prediction of driver behaviour) is some cause for celebration that the method 'works'. This mean that, to some extent at least, 'work as imagined' did indeed match 'work as done'.

The validation performance of the OESDs is comparable with the best of the Human Factors methods (Stanton and Young 1999; Stanton et al. 2013). Some of the best methods in the field include those associated with prediction of task performance time (Card et al. 1983; Baber and Mellor 2001;

Stanton and Baber 2008; Harvey and Stanton 2013; Moray et al. 2017) and prediction of human error (Baber and Stanton 1996; Harris et al. 2005; Stanton and Baber 2002, 2005; Stanton el al. 2009). Whilst a wide range in the predictive performance of Human Factors methods has been observed (Stanton and Young 1999), those better performing methods typically achieve validity statistics in excess of 0.8 (which in correlational terms is generally considered to be of strong predictive value). OESDs appear to be performing as well as could be expected, for the limited range of automation handover tasks, both in driving simulators and a real vehicle on the road.

Whilst the validation evidence for the practical application of OESD is growing, we cannot assume validity generalisation until more extensive studies have been undertaken. This means that it is an important goal for researchers to undertake studies with larger numbers of participants and in a broader range of applications and domains. So far studies have been limited to automation-driver handover tasks and, with the exception of this latest on-road study, mainly in driving simulators (Stanton et al. 2021). As an engineering discipline, Human Factors needs be able to provide evidence that its methods actually work (Stanton and Young 2003; Stanton et al. 2013; Stanton 2016). This means providing evidence for the validity and reliability of the methods to support their continued use.

There is little by way of studies of reliability for OESDs, which is a shortcoming in the research. Reliability has two facets (Stanton and Young 1999; Stanton 2016), stability over time (called intra-analyst reliability) and stability between people conducting the analysis (called inter-analyst reliability). Establishing both forms of reliability is an important future goal for the methodological research into OESD. The repeatability of a method, both within and between analysts, would give confidence in it continued use, although it should be noted that OESDs have been in continuous use for over 60 years without this evidence.

Also little has been written on training people to use Human Factors methods (including OESDs), as noted by Stanton and Stevenage (1998). Stanton and Young (1999, 2003) report on an evaluation of the ease with which novices were able to acquire expertise in Human Factors methods. They should quite a large spread of training times which is, to some extent at least, associated with the complexity of the methods tested. No such study has been reported on the ease or difficulty in acquiring expertise in developing OESD. This represents a gap in knowledge about Human Factors methods, although there are estimates reported in Stanton et al (2013). Typically, the most popular Human Factors methods are those that are quick to learn and apply (Stanton et al. 2013). More research could be undertaken to streamline methods and improve their ease-of-use. Future research could also investigate using the predictive power of OESDs to model performance variability affected by human factor issues such as underload, fatigue or overreliance. This might involve developing a variety of models for the same task to compare with attentive drivers who are not suffering from the effects of underload, fatigue or overreliance. The benefit of such models could be to help the designers of such systems anticipate potential driver maladies and design was to alleviate the symptoms or prevent them from occurring in the first place.

There were some limitations to this study, which are worth noting in terms of full disclosure. First, the sample size was very limited for claims about predictive validity of a method, although the Matthews (1975) Correlation Coefficient was chosen as an acknowledgement of the sampling problem. Second, the modelling of the cognitive activities was problematic. Whilst it may be useful to disassociate the reception of visual-auditory messages from the responses to them for the purposes of system design, it was it was difficult to disentangle the reception of information from the answering of questions in the analysis of the videos. This may have led to an artifactual inflation of Phi, as every cognitive activity would be accompanied by the associated physical activity. Nevertheless, we do consider it important to include both cognitive and physical activities on OESD. Further methodological work is required on the approach. Third, the scope of the study was limited to the handover from vehicle automation to human driver. A wider range of task and domains is required to determine validity generalisation. We are not claiming that the OESDs presented in this study can be generalised to all SAE level 3 driving, rather we are arguing that OESDs can provide a good model, and prediction, of driver behaviour. Finally, as discussed previously, reliability of the method should be established over time as well as between analysts.

6 Conclusions

OESDs are able to predict handover behaviour from vehicle automation to human driver in on-road trials. Although the sample size for this study was quite small compared to the previous driving simulator studies, the findings are encouraging and a large sample size for future work is to be encouraged. For now, the use of OESDs for developing design of the interfaces between vehicle automation and human drivers seems to be reasonable. An important goal for future research is to establish the reliability of OESD generation, both within and between analysts, as well as validity generalisation across a broader range of domains and application.

References

AgeLab. (2019). MIT AgeLab Study Tools. https://agelab.mit.edu/study-tools. Accessed 17 Feb 2020

Baber C, Stanton NA (1996) Human error identification techniques applied to public technology: predictions compared with observed use. Appl Ergonomics 27(2):119–131

Baber C, Mellor B (2001) Using critical path analysis to model multimodal human–computer interaction. Int J Hum Comput Stud 54(4):613–636

Banks VA, Stanton NA, Harvey C (2014) Sub-systems on the road to vehicle automation: hands and feet free but not 'mind' free driving. Saf Sci 62:505–514

Banks VA, Plant KL, Stanton NA (2018a) Driver error or designer error: using the perceptual cycle model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. Saf Sci 108:278–285

Banks VA, Stanton NA, Burnett G, Hermawati S (2018b) Distributed cognition on the road: using EAST to explore future road transportation systems. Appl Ergon 68:258–266

Bysameyee (2020) Head-up display GPS navigation. https://www.amazon.co.uk/Head-up-Display-GPSNavigation-Reflective/dp/B075T2DH12/ref=sr_1_10?keywords=car+head+up+display&qid=1573804882&sr=8-10. Accessed 17 Feb 2020

Card SK, Moran TP, Newell A (1983) The psychology of human-computer interaction. Lawrence Erlbaum, Hillsdale

Clark J, Stanton NA, Revell KMA (2019a) Distractibility, eye-gaze, and the usage of visual displays during an automated vehicle handover task. Transp Res Part F: Psychol Behav 67:29–42

Clark JR, Stanton NA, Revell MA (2019b) Identified handover tools and techniques in high-risk domains: using distributed situation awareness theory to inform current practices. Saf Sci 118:915–924

Eriksson A, Stanton NA (2017a) Take-over time in highly automated vehicles: transitions to and from manual control. Hum Factors 59(4):689–705

Eriksson A, Stanton NA (2017b) Driving performance after self-regulated control transitions in highly automated vehicles. Hum Factors 59(8):1233–1248

Eriksson A, Petermeijer SM, Zimmerman M, De Winter JCF, Bengler KJ, Stanton NA (2019) Rolling out the red (and green) carpet: supporting driver decision making in automation to manual transitions. IEEE Trans Hum Mach Syst 49(1):20–31

Green DM, Swets JA (1966) Signal detection theory and psychophysics, vol 1. Wiley, New York

Harris D, Stanton NA, Marshall A, Young MS, Demagalski J, Salmon P (2005) Using SHERPA to predict design-induced error on the flight deck. Aerosp Sci Technol 9:525–532

Harris D, Stanton NA, Starr A (2015) Spot the difference: Operational event sequence diagrams as a formal method for work allocation in the development of single-pilot operations for commercial aircraft. Ergonomics 58(11):1773–1791

Harvey C, Stanton NA (2013) Modelling the hare and the tortoise: predicting the range of in-vehicle task times using critical path analysis. Ergonomics 56(1):16–33

Hollnagel E (2015) Why is work-as-imagined different from work-as-done? In: Wears RL, Hollnagel E, Braithwaite J (eds) Resilient health care—the resilience of everyday clinical work, vol 2. Aldershot, Ashgate, pp 249–264

Kirwan B, Ainsworth LK (eds) (1992) A guide to task analysis: the task analysis working group. CRC Press, Boca Raton

Kurke MI (1961) Operational sequence diagrams in systems design. Hum Factors 3:66–73

Landis JR, Kock GG (1977) The measurement of observer agreement for categorical date. Biometrics 33:159–174

Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta BBA Protein Struct 405(2):442–451

Moray N, Groeger J, Stanton NA (2017) Quantitative modelling in cognitive ergonomics: predicting signals passed at danger. Ergonomics 60(2):206–220

Read GJM, Stevens EL, Lenné MG, Stanton NA, Walker GH, Salmon PM (2018) Walking the talk: comparing pedestrian 'activity as imagined' with 'activity as done.' Accid Anal Prev 113:74–84

SAE On-road Automated Vehicles Standards Committee. (2016). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Accessible: http://standards.sae.org/j3016_201401/

Salmon PM, Stanton NA, Walker GH, Jenkins DP, Baber C, McMaster R (2008) Representing situation awareness in collaborative systems: a case study in the energy distribution domain. Ergonomics 51(3):367–384

Salmon PM, Walker GH, Stanton NA (2016) Pilot error versus socio-technical systems failure: a distributed situation awareness analysis of Air France 447. Theor Issues Ergon Sci 17(1):64–79

Sarter NB, Woods DD (1995) How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum Factors 37(1):5–19

Sorensen LJ, Stanton NA, Banks AP (2011) Back to SA school: contrasting three approaches to situation awareness in the cockpit. Theor Issues Ergon Sci 12(6):451–471

Stanton NA (2016) On the reliability and validity of, and training in, ergonomics methods: a challenge revisited. Theor Issues Ergon Sci 17(4):345–353

Stanton NA, Baber C (2002) Error by design: methods for predicting device usability. Des Stud 23:363–384

Stanton NA, Baber C (2005) Validating task analysis for error identification: reliability and validity of a human error prediction technique. Ergonomics 48(9):1097–1113

Stanton NA, Baber C (2008) Modelling of alarm handling responses times: a case of the Ladbroke Grove rail accident in the UK. Ergonomics 51(4):423–440

Stanton NA, Marsden P (1996) From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. Saf Sci 24(1):35–49

Stanton NA, Stevenage S (1998) Learning to predict human error: issues of reliability, validity and acceptability. Ergonomics 41(11):1737–1756

Stanton NA, Young M (1999) What price ergonomics? Nature 399:197–198

Stanton NA, Young MS (2003) Giving ergonomics away? The application of ergonomics methods by novices. Appl Ergon 34:479–490

Stanton NA, Young M, McCaulder B (1997) Drive-by-wire: the case of mental workload and the ability of the driver to reclaim control. Saf Sci 27(2–3):149–159

Stanton NA, Salmon P, Harris D, Marshall A, Demagalski J, Young MS, Waldmann T, Dekker SWA (2009) Predicting pilot error: testing a new methodology and a multi-methods and analysts approach. Appl Ergon 40(3):464–471

Stanton NA, Dunoyer A, Leatherland A (2011) Detection of new in-path targets by drivers using stop, & go adaptive cruise control. Appl Ergon 42(4):592–601

Stanton NA, Salmon PM, Rafferty LA, Walker GH, Baber C, Jenkins D (2013) Human factors methods: a practical guide for engineering and design, 2nd edn. Aldershot, Ashgate

Stanton NA, Salmon PM, Walker GH, Salas E, Hancock PA (2017) State-of-science: Situation awareness in individuals, teams and systems. Ergonomics 60(4):449–466

Stanton NA, Salmon PM, Walker GH, Stanton M (2019) Models and methods for collision analysis: a comparison study based on the Uber collision with a pedestrian. Saf Sci 120:117–128

Stanton NA, Brown J, Revell KMA, Langdon P, Bradley M, Politis I, Skrypchuk L, Thompson S, Mouzakitis M (2021) Validating Operator Event Sequence Diagrams: the case of automated vehicle to human driver handovers. Hum Factors Ergonomics Manuf Serv Ind. https://doi.org/10.1002/hfm.20887

Walker GH, Stanton NA, Baber C, Wells L, Gibson H, Salmon PM, Jenkins DP (2010) From ethnography to the EAST method: a tractable approach for representing distributed cognition in air traffic control. Ergonomics 53(2):184–197