

Validating Operator Event Sequence Diagrams: The case of an automated vehicle to human driver handovers

Neville A. Stanton¹  | James Brown¹ | Kirsten M. A. Revell¹ | Pat Langdon² | Mike Bradley² | Ioannis Politis² | Lee Skrypchuk³ | Simon Thompson³ | Alexandros Mouzakitis³

¹Transportation Research Group, Civil, Maritime, Environmental Engineering and Science, Faculty of Engineering and the Environment, University of Southampton, Southampton, UK

²Engineering Design Centre, Engineering Department, University of Cambridge, Cambridge, UK

³Jaguar Land Rover, Research & Technology, Coventry, UK

Correspondence

Neville A. Stanton, Transportation Research Group, Civil, Maritime, Environmental Engineering and Science, Faculty of Engineering and the Environment, University of Southampton, Boldrewood Innovation Campus, Burgess Rd, Southampton SO16 7QF, UK.

Email: ns4c08@soton.ac.uk

Funding information

Jaguar Land Rover and the Engineering and Physical Sciences Research Council, Grant/Award Number: EP/N011899/1

Abstract

Predicting what drivers will do as vehicle control is handed over to them from automation is a relatively new challenge for the motor vehicle industry. Operator Event Sequence Diagrams (OESDs) offer a way of modeling the interactions between the driver and vehicle automation in the handover of control. In this paper, two studies are presented in which a range of handover strategies are tested. The anticipated driver strategies were modeled using OESDs to serve as predictions of driver behavior. Drivers were then observed in two separate studies: (1) using a Lower-Fidelity (vehicle seat and controls) simulator and (2) using a Higher-Fidelity (whole vehicle) simulator. Driver behavior during a takeover task was categorized according to the signal detection paradigm into hits, misses, false alarms, and correct rejections. The results showed that for all strategies in both sets of studies, the median criterion for validity was exceeded ($\varphi > 0.8$), suggesting that OESDs made good predictions of driver behavior during the handover of the vehicle from automation to manual control.

KEYWORDS

automation, driving, modeling, OESD, validation

1 | INTRODUCTION

With the impending arrival of highly automated vehicles (Stanton et al., 2020), the handover of control from computing technology to human drivers becomes an important issue (Eriksson & Stanton, 2017). This issue may be exacerbated if the driver has not been monitoring the vehicle and the road environment, as defined in SAE level 3 automation (SAE, 2018: J3016). It is anticipated that level 3 automation will enable drivers to undertake nondriving activities, while the computers drive the vehicle for them. Such nondriving activities could include reading, checking emails,

surfing the internet, watching movies, making phone or video calls, eating and drinking, or even playing video games (Stanton, 2015). For the purposes of our research, we envision driving automation on motorways (freeways) only and that the driver would remain the fall-back for planned handovers (i.e., nonemergency conditions) when the automated system reaches operational limits (such as approaching a motorway/freeway exit). Bringing the driver back into control of the vehicle after some time of nonengagement is not a trivial undertaking. This will require drivers to become aware of a range of situational and contextual features, including (but not limited to): the road environment, weather conditions, behavior

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Human Factors and Ergonomics in Manufacturing & Service Industries* Published by Wiley Periodicals LLC

of other vehicles, hazards on the road, status of their own vehicle, as well as immediate actions required, such as vehicle navigation and guidance (Stanton et al., 2017; Walker et al., 2015).

The design of the handover interface requires some form of representation of the human and machine activities that are anticipated so that the strategy can be implemented in the vehicle. There is a range of Human Factors and Ergonomics methods that could be employed to represent these activities (Stanton et al., 2013, 2014). Modeling of the interaction between people and technology is becoming increasingly popular in Human Factors and Ergonomics research and practice (Moray et al., 2016). These methods enable the modeling of tasks, processes, timings, and potential errors. Choosing the most appropriate methods really comes down to the intended purpose to which the modeling is to be put (Stanton et al., 2013). Modeling the structure of tasks can be useful for understanding the nature of work design (Stanton, 2006). Modeling processes helps in the design of interaction between humans and machines (Banks & Stanton, 2017). Modeling timings can be used to understand time-critical interventions (Stanton & Baber, 2008; Stanton & Walker, 2011). Modeling errors can be helpful for anticipating non-normative behavior (Stanton et al., 2009). For the purposes of modeling the process of handover from an automated vehicle to a human driver, Operator Event Sequence Diagrams (OESDs) were selected as they have

been successfully used to model the interactions between drivers and vehicles previously (Banks & Stanton, 2017; Banks et al., 2014).

OESDs have been used to represent the interaction between humans and technology in a graphical manner (Stanton et al., 2013). They are based on the engineering technique for describing technical processes in the form of a flowchart with the use of standard symbols for each defined process (see Table 1). Each separate system has its own column, colloquially called “swim-lanes.” OESDs have additional “swim-lanes” for the human operator(s). The output of an OESD graphically depicts activities, including the tasks performed and the interaction between humans and machines over time, using standardized symbols. There are numerous forms of OESDs, ranging from a simple flow diagram representing task order to more complex OESD, which account for team interaction and communication and often include a timeline of the scenario under analysis and potential sources of error.

Previous applications of OESDs include modeling single-pilot operations in commercial aviation (Harris et al., 2015), aircraft landing (Sorensen et al., 2011), air traffic control (Walker et al., 2010), electrical energy distribution (Salmon et al., 2008), and automatic emergency braking systems in automobiles (Banks et al., 2014). The latter of these applications used OESDs to compare four different levels of automation for pedestrian detection and avoidance from manual control, through

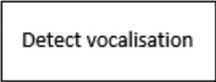
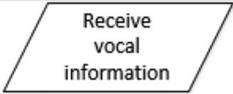
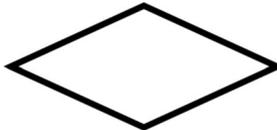
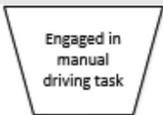
| Example | Description | OESD elements |
|---|---|--|
|  | A generic task |  |
|  | Information is received |  |
|  | A decision with two or more possibilities |  |
|  | Audio output |  |
|  | Display output |  |
|  | A manual task |  |

TABLE 1 OESD task elements with description and an example

Abbreviation: OESD, Operator Event Sequence Diagram.

decision support and automated decision-making to full automation. This analysis showed that both the decision support and automated decision-making systems actually involved the driver in more work than driving manually. All of the applications of OESDs were able to delineate between the processes undertaken by different human actors and machine agents in their respective systems. For example, Harris et al. (2015) showed the effects of reducing the crew of two pilots to a single pilot. Single-pilot operations reduced many of the crew communication activities but did result in more work overall for the remaining pilot, as might be expected.

In the other applications of OESDs, researchers have been able to show how work is distributed across multiple actors and agents. In particular, Sorensen et al. (2011) used OESDs to illustrate how distributed cognition (Hutchins, 1995a, 1995b) worked in their analysis of an aircraft cockpit and crew for a landing task. Inspired by the original work of Edwin Hutchins (1995b), they used OESDs to show how the cognition of the cockpit is distributed among the artifacts, and two human pilots as the descent of the aircraft are managed. Similarly, Walker et al. (2010) and Salmon et al. (2008) show how cognition is distributed among the artifacts and people (who are themselves in different physical locations) for aviation and energy distribution industries, respectively. In a recent review of distributed cognition and situation awareness (SA), Stanton et al. (2017) argue that automated driving systems provide an excellent case study for distributed cognition. Clearly, the vehicle automation is performing some of the cognitive functions on behalf of the human driver, the extent to which depends upon the level of automation involved. OESDs can make the distribution of these cognitive functions more explicit, as well as identifying the interactions between the human driver and automated systems when conducting the handover (Banks et al., 2014).

While modeling of the interaction may be useful for designing the strategy for the handover, it is only a prediction of the behavior of the system. As such, it requires validation. Unfortunately, validation evidence is rarely reported in the literature (Stanton, 2014, 2016; Stanton & Young, 1999a, 1999b). There are a few notable exceptions, however, such as the development of the human error prediction methods (Stanton & Baber, 2005; Stanton & Stevenage, 1998; Stanton et al., 2009). Stanton et al. (2013) report the state of evidence of reliability and validity for 109 methods, which changed little since their first review in 2005. Stanton (2016) has argued strongly for reporting this evidence when using Human Factors and Ergonomics methods. In addition, even when there is such evidence for the use of a method in one domain, validity generalization to other domains cannot be assumed and requires testing.

Previous validation studies of Human Factors and Ergonomics methods have used the signal detection paradigm (Stanton & Young, 1999a, 1999b). This approach has enabled analysts to distinguish between predictions of behaviors that are observed (the hit rate) and those that are not (the false alarm rate). It is also possible to calculate the overall sensitivity of the methods, taking both the hit rate and the false alarm rate into account. In summary, the purpose of this paper is to develop OESDs for the vehicle to driver (VtD) handover process and then to validate them in studies using driving simulators. It was

anticipated that the models of handover would offer good predictions of actual handover paper, based on modeling evidence from other methods and domains (e.g., Stanton et al., 2013, 2014).

2 | OESD DEVELOPMENT

A set of interaction concepts to facilitate VtD handover were derived following two design workshops. The workshops comprised psychologists, human factors engineers, computer scientists, and automotive engineers. During the course of the workshops, discussions were held on the role of the driver and the role of vehicle automation. These discussions were refined into storyboards that revealed the processes, tasks, and agents involved. From the storyboards, the OESD swimlanes were drawn up. This was an iterative process, entailing drawing up the tasks and processes in the OESDs, comparing with the storyboards, and discussing with the research team. OESDs were selected for their ability to illustrate the tasks and processes with respect to time (Banks et al., 2014; Harris et al., 2015; Sorensen et al., 2011). Events that occur within the system are modeled using a set of task unit elements, the type of element depends on the event type. Table 1 illustrates the set of task unit elements that correspond to different events. Time is represented on the vertical axis of OESDs; therefore, two events occurring simultaneously would be at the same level, whereas sequential events would be vertically offset.

Four interaction designs were selected from the set of concepts, and each represented using an OESD. Each OESD modeled the interactions between the agents throughout a complete cycle of the automation system. The cycle started with the driver in manual mode undertaking a driver to vehicle handover (DtV). During the automated period (when the vehicle is being driven by automation), the OESD modeled the driver interacting with a secondary task (which was engaging in a tablet-based memory game) followed by the VtD handover that returned the driver to manual control.

Figure 1 illustrates a part of an OESD, the four agents, in this case, being driver, vehicle, handover interaction, and environment. Phase A represents the start of the handover of control, where the driver is informed of a forthcoming mode change, and they begin to prepare. The vehicle is currently automated, however in processes not shown in the diagram, the vehicle's sensors have detected the need for the driver to take control, and this has resulted in the vehicle making the handover interaction vocalization. The six driver processes of Phase A then follow. The driver immediately receives the information then performs a sequence of manual actions, ceasing interaction with a secondary task device, putting down the secondary task device, resuming a driving position, and paying attention to the road. The driver then vocalizes that they are ready to take control of the vehicle, followed by the vehicle detecting the vocalization. In the event that the driver makes an unintelligible or negative response, the initial vocalization is repeated.

Phase B, which contains the handover protocol processes, then begins with the vehicle generating hazard information from navigational data created by on-board sensors (see Politis et al., 2018).

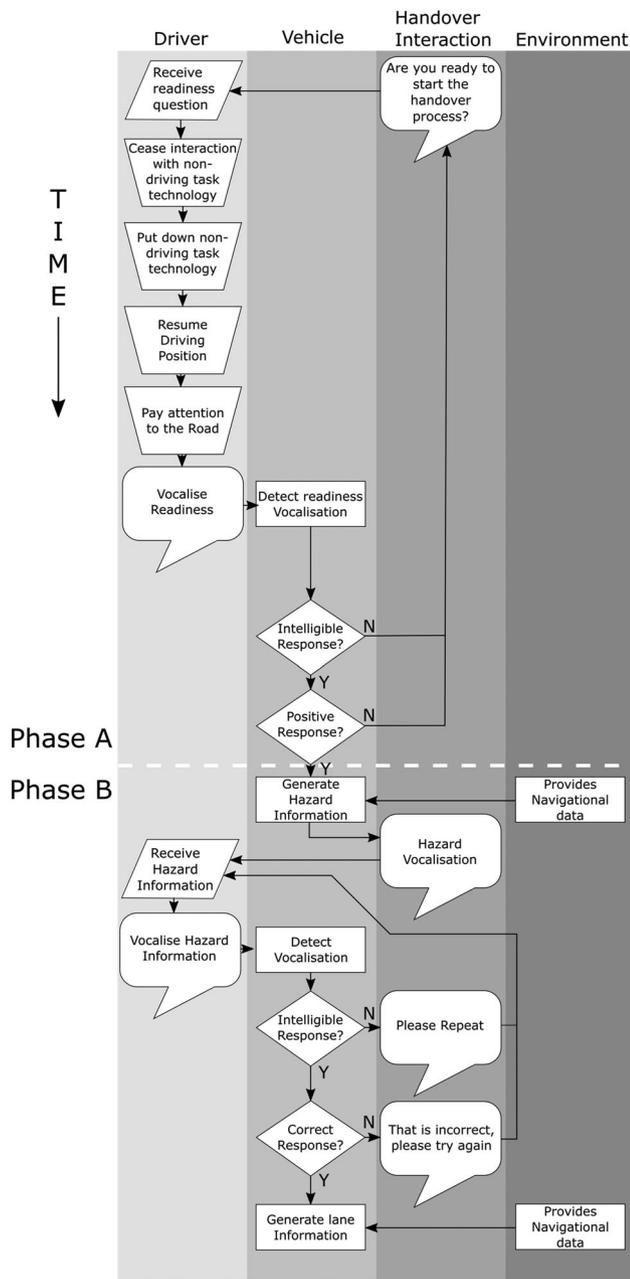


FIGURE 1 Example OESD section showing all six driver processes for Phase A (Start of VtD) and the first two driver processes of Phase B (Handover protocol phase). This OESD is an example from Study 2 (Higher-Fidelity Simulator study). OESD, Operator Event Sequence Diagram

The vehicle vocalizes the hazard information, which results in the first two driver processes of Phase B; the driver receives the hazard information and then vocalizes it. In the event that the vocalization is unintelligible, the driver is asked to repeat. If the driver answers incorrectly, they are informed of their incorrect answer and asked to try again. Once the driver has provided the correct response, the next part of the protocol starts, in this case, lane information.

To determine how well handover behavior was modeled, simulator studies were conducted in both a Lower-Fidelity (Study 1: Politis

et al., 2018) and Higher-Fidelity (Study 2) simulators. These two studies were undertaken as part of the handover design process. The handover concept first being tested in the lower fidelity simulator and then again in the higher fidelity simulator. After the completion of the Lower-Fidelity experiments, changes were made to the interaction designs to provide a better driver experience. The OESDs were amended to reflect these changes allowing the method to be further tested in the Higher-Fidelity experiment. This double testing also allowed the OESDs to be validated in both simulators to see if the fidelity of the driving simulation environment made any difference.

3 | STUDY 1—VALIDATION OF OESD MODELED DRIVER BEHAVIOR IN A LOWER-FIDELITY DRIVING SIMULATOR

3.1 | Participants

Forty-nine participants took part, 24 of whom were female, ages ranged from 17 to 86 years (mean = 45.51, $SD = 17.36$). All participants held a full driving license for between 1 and 70 years. (mean = 25.68, $SD = 17.58$). An approximate gender split between age ranges was achieved (7 m, 5 ft, 17–34), (11 m, 13 ft, 35–56), (7 m, 6 ft, 58–86). This was in-line with the principles of inclusive design (Keates & Clarkson, 2003). Ethical approval was granted by the Ethics Review Committee, University of Cambridge, Department of Engineering.

3.2 | Experiment design

The experiment employed a repeated measures design with four VtD control transfer conditions representing the single independent variable. Conditions were counterbalanced using the Latin Square design. All conditions featured an initial stage whereby the system would verbally ask if the driver was ready to resume control, following a verbal confirmation, the protocol would start. The first condition (Timer) was based on a simple timer that appeared on the dashboard display when the automation detected a need to hand control back to the driver; on confirmation that the driver was ready, it counted down in 10 s intervals from 60 s, the driver is required to take control by pressing a button on the steering wheel before the countdown reached zero. This was based on an existing design currently undergoing testing by Volvo (Volvo Cars, 2015) with an auditory rather than visual countdown.

The second condition, “HazLan” used a “readback” to raise SA, in which the system would vocalize five elements of SA: potential hazards, current lane, current speed, the next required exit, and the next required action. Following each of the system’s vocalizations, the participant was required to repeat back. Incorrect or missing readbacks resulted in the system repeating the original vocalization. Once all of the readbacks were complete, the participant was able to resume manual control.

The third condition, "VAA" was response-based using the same element sequence as the second condition; however, the system provided the participants with a question regarding each element. If the participant answered the question correctly, the next question was presented; upon completion of the sequence, the participant was able to resume manual control.

The fourth condition was an augmented version of the HazLan condition. In addition to the HazLan SA aspect sequence, it incorporated multiple modalities in the form of audio-driven seat-based haptics (whereby audio signals, including vocalizations, were transmitted to the driver via a pad on the driver's seat) and two light-emitting diode (LED) strips mounted on each side of the driving position. This presented constant information on the longitudinal positions of cars in neighboring lanes, thus providing a dynamic blind-spot warning system.

Data on multiple aspects were collected; however, for the validation experiment presented in this paper, the dependent variable was driver behavior, in terms of the processes carried out by participants during VtD control transitions. This driver behavior data was collected using four webcams, generating footage from multiple angles within the vehicle. The processes consisted of actions, inactions, and vocalizations. Actions included those expected from the driver as predicted by the associated OESD, as well as unexpected actions, such as placing a finger on a button early or making an exaggerated glance. Inactions consisted of the failure to carry out a process predicted by the associated OESD. Vocalizations included those expected by the system as specified in the associated OESD.

3.3 | Equipment

Experiments were carried out using a lower-fidelity driving simulator consisting of a gaming seat, a Logitech G25 steering wheel and pedal set, and three screens to provide a wide field of view (as shown in Figure 2). An additional tablet was employed to act as a pseudo-dashboard to illustrate speed, lane positioning relative to other cars, fuel level, automation mode, ideal lane, and the next required exit.



FIGURE 2 Study 1 (Lower-Fidelity) experimental configuration (Politis et al., 2018)

The driving scenario featured a route approximately 10 miles long and consisting of a combination of highways with gentle bends and urban roads without corners. A Java-based memory game application was installed on a tablet-based PC to provide the participant with a secondary task when automation was enabled. Participant behavior was recorded using a camera with a wide-angle lens. Two Arduino-based LED strips were fitted to the wall on each side of the driving position, C code was written to enable them to perform as blind spot indicators. A TAD (tactile acoustic device) was placed on the driving seat to provide sound-based haptic information.

3.4 | Procedure

After welcoming, the participants were briefed on the experiment and presented with a demographics questionnaire. The simulator was then presented to the participants, and they embarked on a short introductory test drive. No other information was provided to the participants, other than a brief overview of the vocalization system, to avoid training effects. The driving scenarios were then run, using a counterbalanced design to mitigate order effects. During each automation phase, participants were requested to play the tablet-based memory game; a total of three VtD transitions were performed per scenario, at approximately 10%, 40%, and 70% progress through the route. VtD transition dialogues used synthetic vocalizations; the participants were expected to respond vocally before switching to manual control using a button on the steering wheel. A Wizard of Oz-based system was employed to manage the synthetic vocalizations in response to the participant. At the end of the study, participants were provided with remuneration in the form of a £20 web voucher for their time.

3.5 | Analysis

Signal detection theory's primary use is to discern between "signals" and noise (Abdi, 2007), four stimulus-response events exist: Hits, Misses, False Alarms (FAs), and Correct Rejections (CRs) (Nevin, 1969). In the context of this experiment, it provided a method by which to compare participant behavior, observed during experiments, with predicted driver behavior illustrated on OESDs. See Figure 3.

For the Lower-Fidelity experiments, the analysis was limited to the second control transition for each condition; three conditions were analyzed: Timer, HazLan, and VAA.

Labeled template forms consisting of SDT matrices were created to allow the paper-based recording of the analysis. The footage was displayed on a large liquid-crystal display (LCD) screen and printed sequential lists of VtD driver processes for each condition drawn from the OESDs were used for reference. The footage of the selected VtD transfers was viewed, and participants' behaviors were noted on the SDT matrices using the OESD derived driver processes as a template of expected behavior. To

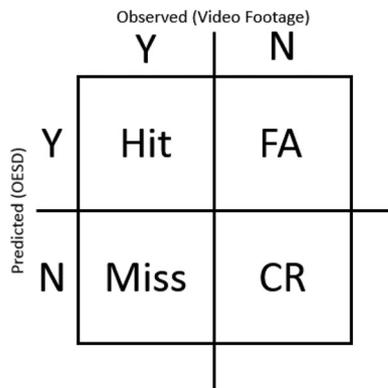


FIGURE 3 Signal detection theory (SDT) matrix. CR, correct rejection; FA, false alarm; OESD, Operator Event Sequence Diagram

aid the analysis, the driver processes were split into three phases, A, B, and C, representing the participant preparing to take control, proceeding through the protocol, and taking control back from the automation, respectively. The Timer condition was particularly short and therefore had no requirement for Phase B. Figure 1 shows the six driver-based processes of Phase A in the driver column.

A “perfect” SDT score was attained when the participant only carried out every predicted process as part of the VtD control transfer; in this case, the SDT matrix would have an equal number of “Hits” as there were predicted driver processes. For each predicted process that a participant failed to carry out, a “False Alarm” was recorded. In the event that a driver exhibited behavior in addition to that which was predicted, a “Miss” was recorded. CRs were calculated at the end of the SDT analysis by subtracting the number of FAs from the total pool of all FAs for all participants. In summary, the data were processed as follows.

3.5.1 | Hits

Present in OESD and present in the video of automation-driver handover (e.g., the driver vocalizing their readiness to begin the handover process).

3.5.2 | Misses

Not present in the OESD but present in the video (e.g., the driver resuming control of the vehicle early and not progressing through all of the handover protocol).

3.5.3 | False alarms

Present in the OESD but not present in video (e.g., the driver does not vocalize hazard information).

3.5.4 | Correct rejections

Not present in OESD 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 the participants, minus the number of misses for each individual participant.

The results from all the SDT matrices were collated in a spreadsheet, allowing ϕ to be calculated. The Matthews correlation coefficient (ϕ) was applied to the data generated by the SDT analysis, this quantified the correlation between the expected and observed behavior as a means to validate the OESDs. The Matthews correlation coefficient formula is shown below:

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

3.6 | Inter-Rater Reliability (IRR) Method

IRR testing was carried out due to the subjective nature of analyzing, interpreting, and categorizing driver behavior. An analyst was provided with approximately 10% of the video footage files, together with associated SDT analysis forms, a list of exceptions, and a list of driver processes split across the three phases. The analyst watched the footage and compared the driver behavior to that which was expected as specified in the list of driver processes. SDT results were recorded on the SDT analysis forms, together with any exceptions that occurred. This was identical to the method used by the original analysts. Equal weighted Cohen's κ values were calculated and are reported in Section 3.7.

3.7 | Results

Equal weighted Cohen's κ values were calculated, resulting in a value of 0.773 for the Lower-Fidelity simulator. This represents a moderate agreement between the analysis in their classification of hits, misses, FAs, and CRs (Landis & Koch, 1977).

As shown in Figure 4, all three experimental conditions exhibited a relatively high number of hits per condition, and all shared identical interquartile ranges of 1. The Timer condition did not require verbal interaction with the driver, resulting in fewer possible hits than the Hazlan and VAA conditions. All experimental conditions contained outliers.

In terms of misses by condition, both the HazLan and VAA conditions had identical median values (2) and interquartile ranges (3). The shorter interaction steps in the Timer condition may have contributed to the lower median value of 1 and interquartile range of 2.

In terms of FAs, all conditions showed equal interquartile ranges of 1 and shared identical outlier values. Median values for the HazLan and VAA conditions were also equal at 1. The VAA condition exhibited a slightly higher number of CRs (16) than the HazLan (12) and Timer (13) conditions; all interquartile ranges were identical (1).

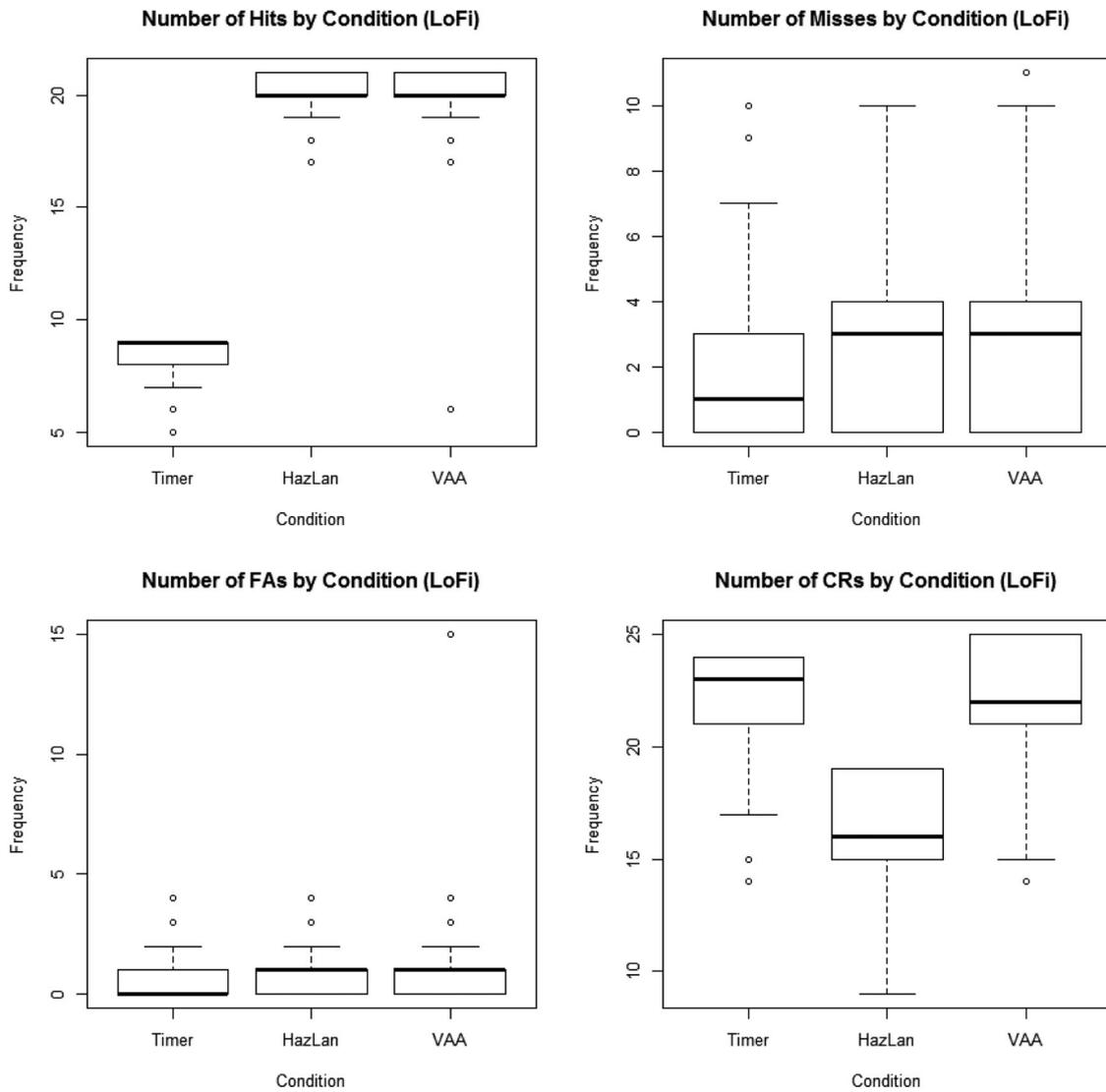


FIGURE 4 Lower-Fidelity experiment hits, misses, FA, and CR by condition. CR, correct rejection; FA, false alarm

As shown in Figure 5, all median values for Matthews correlation coefficient (ϕ) by condition were greater than 0.8 (the minimum acceptable criterion), indicating a strong positive relationship. The Timer condition scored particularly high with a value over 0.9. The Timer condition also exhibited a large interquartile range, varying from slightly above 0.7 to 1.0. The HazLan condition interquartile range was slightly lower at approximately 0.82, the smallest interquartile range was found in the VAA condition at approximately 0.15.

4 | STUDY 2—VALIDATION OF OESD- MODELED DRIVER BEHAVIOR IN A HIGHER-FIDELITY DRIVING SIMULATOR

4.1 | Participants

A total of 60 participants were recruited (30 males, 30 females) with an equal gender split across three age ranges, 18–34 (mean = 25.8,

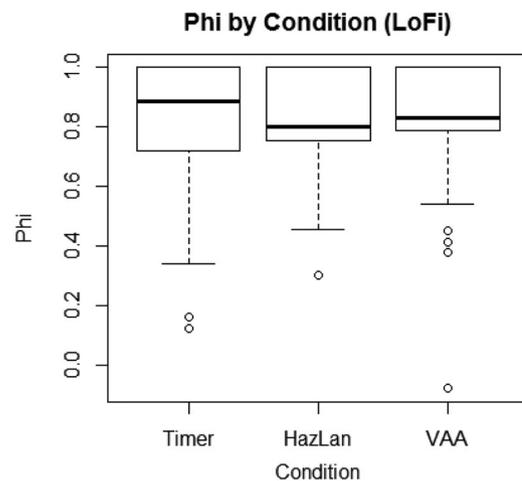


FIGURE 5 Study 1 Lower-Fidelity Experiment Phi (ϕ) values by condition

$SD = 4.2$), 35–56 (mean = 44.2, $SD = 7.8$), and 57–82 (mean = 63.2, $SD = 4.8$). All participants were holders of the current U.K. driving licenses and signed the informed consent form. The research was approved by the University of Southampton Ethics Research and Governance Office (ERGO No. 29615).

4.2 | Experiment design

A repeated measures experiment was selected, four different VtD control transfer conditions were tested. The type of handover condition formed the independent variable, each featuring a unique combination of HDD (Head Down Display) format, HUD (Head Up Display) format, and visual or haptic ambient indicators. Participants always experienced the control condition (Timer) first, followed by the remaining three conditions, counterbalanced to mitigate learning and fatigue effects.

The timer was the control condition and consisted of a simulated version of Volvo's timer-based interaction design that depicts a visual countdown (Volvo Cars, 2015). On initiation, the HDD displayed an instruction to resume control, a 60 s countdown timer and an associated graphic updating at 2 s intervals. When the participant took manual control by pressing a button marked "M" on the steering wheel, the HDD icons and automation icon would disappear.

HazLan Condition 1 presented the driver with instructions via synthesized vocalizations and icons on the HDD and HUD. The driver was first asked verbally by the system if they were ready to start the handover process. The protocol then consisted of five elements of information for highway sections regarding Hazards, Lane, Speed, Exit, and Action (HazLanSEA). For urban sections, the Lane information was removed, as the route featured a single carriageway. These elements were presented sequentially, as icons on the HDD and HUD, and as corresponding synthesized vocalizations. The driver was expected to acknowledge that they had received each element of information by repeating it back, at which point the next was presented. After the last element of the protocol was acknowledged, the driver was presented with an icon and verbalization, asking them to resume manual control by pressing the "M" button on the steering wheel.

HazLan condition 2 consisted of an augmented version of HazLan condition 1, with the addition of vehicle highlighting and radar range indicators on the HUD. The radar range was indicated by means of a large green circle at ground level and centralized on the participant's car. The radius of the circle represented the range at which other vehicles were detected. Vehicles were highlighted by placing them within a red-framed box once they were within range.

HazLan condition 3 consisted of HazLan condition 2, further augmented with two ambient systems. A haptic seat base provided a short, strong vibration alert to the participant at the condition initiation. LED strips mounted on the A-pillars provided the participant with continual feedback of the longitudinal position of vehicles in the immediately opposite lanes.

At the end of every road section of conditions 2, 3, and 4, the participant was presented with an icon, accompanied by a synthesized vocalization, asking them to take manual control of the car by pressing the "M" button on the steering wheel. This completed the condition by transferring control from the Vehicle to the Driver (VtD).

The dependent variable for the Higher-Fidelity study was driver behavior, captured via video footage, in an identical way to that of the Lower-Fidelity study. A wide range of additional data relating to performance and the subjective experience was also collected as part of this study but only data relevant to the validation focus of this paper will be discussed.

4.3 | Equipment

The driving simulator consisted of a Land Rover Discovery Sport, in combination with three frontal screens, providing a 140 field of view. Door mirrors employed LCD screens and the rearview mirror reflected a rear-mounted projection. Vehicle control data was extracted from the Controller Area Network using a hardware interface and software in C#. STISim Drive® Version 3 simulator software provided the virtual environment. The vehicle's cluster was replaced by a Microsoft Surface tablet running a custom dashboard in C#; this displayed speed, vehicle position, and icons/graphics (as shown in Figure 6). An HUD was generated by displaying 2D icons and 3D objects in the driver's field of view using STISim's Open Module. An Arduino-based haptic system was fitted to the driver's seat. Arduino-based LED strips were fitted to the A-pillars and coded to act as blind spot indicators. A Microsoft Surface tablet was employed to run a C# Wizard of Oz application, allowing control over HUD and HDD icons graphics, audio vocalizations, and haptics as part of the handover protocol. Four webcams were fitted within the vehicle's interior, the footage from which was saved using a standalone PC. A mini tablet was employed as a secondary task, running a Java-based memory game. All audio outputs were fed into the vehicle's line-in, providing the driver with stereo sound via the internal speakers.

Figure 6a illustrates the view from the driving seat of the Discovery. The tablet-based dashboard can be seen in place of the cluster, and an HUD icon is visible. The LED strip is visible at the base of the right A-pillar. Figure 6b illustrates just the HUD and HDD portions; a protocol icon, radar range, and vehicle highlighting is visible on the HUD. The HDD shows the same icon, speed, vehicle position, and automation status.

4.4 | Procedure

Before the start of the experiment, participants were welcomed, briefed on health and safety issues, and provided with a basic overview of the experiment and hardware before being presented with a consent form, patient information sheet, and biographical

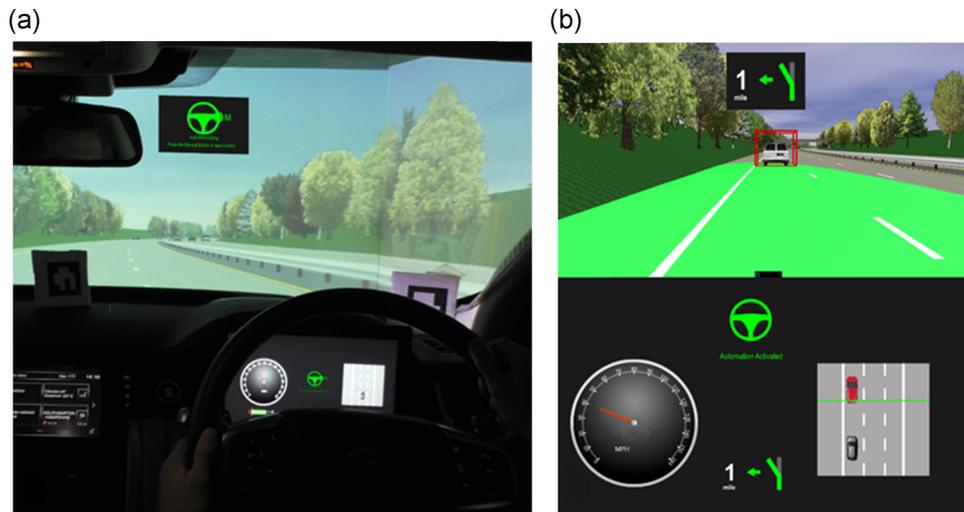


FIGURE 6 Study 2 (Higher-Fidelity) experimental configuration

sheet. The principles of SAE level 3 automation were explained, together with the purpose of the study and the forms of technology in each of the four conditions. Participants were shown a map of the route, indicating the placement of roundabouts necessitating transition of control back to manual, and they were also informed that GPS-style verbal directions would be supplied. The tablet-based secondary task was explained, and they were instructed to use it when the automation mode was enabled. The participants were then asked to enter the vehicle before being introduced to the systems.

A custom route was created for the Higher-Fidelity study; this comprised two lengths of a highway, separated by sliproads to roundabouts, followed by two urban stretches, separated by roundabouts. The route took approximately 20–25 min to negotiate when obeying U.K. traffic laws. Each highway section took approximately 5 min to complete, and each urban section 2.5 min. Road conditions were consistently dry and well lit with a traffic density of approximately one vehicle per 500 ft in all lanes.

At the start of each condition, participants were instructed via a synthesized audio voice to drive manually onto the route; they were then informed that automation was available via a synthesized vocalization and icon. The participant activated automated mode by pressing a button marked “A” transferring control from the DtV. Participants then picked up and interacted with a secondary task in the form of a tablet-based memory game. At a prescribed point on the route, the VtD handover interaction design specific to the experimental condition was activated by the experimenter using a Wizard of Oz (WoZ) interface. This process occurred twice on the highway and twice on the urban section of the route.

For HazLan conditions 1–3, the participant was expected to verbally acknowledge and prepare to start the protocol. From this point, the driver was presented with a sequence of SA-based information in sync with the dynamic context of the road and car status. Information was displayed using a combination of HDD, HUD,

acoustic, and ambient systems, with the participant acknowledging through verbal readbacks. When all elements of the protocol were acknowledged, the experimenter used the WoZ interface to ask the participant to regain manual control of the vehicle. For the Timer condition, information was presented to the driver via the HDD, which presented a simple circular dial-based 60 s countdown and automation mode icon on the HDD. Neither audio cues were provided nor readback confirmations required. The participant was expected to respond to the request to take control before the 60 s had passed. On completion of each condition, participants were asked to fill out the questionnaires; at the end of the fourth condition, they completed a freeform questionnaire and received remuneration in the form of £10 to cover their travel costs.

4.5 | Analysis

Analysis of the Higher-Fidelity experimental data was identical to that described for the Lower-Fidelity in Study 1.

4.6 | Results

Equal weighted Cohen's κ values were calculated, resulting in a value of 0.819 for the Higher-Fidelity simulator. This indicates good agreement between analysts in the classification of hits, misses, FAs, and CRs (Landis & Koch, 1977).

Results from the SDT analysis were used to generate box plot graphs in R. Figure 7 illustrates the hit/miss/FA and CR outputs of the SDT matrices ordered by the condition.

All HazLan conditions exhibited similar median values for the number of hits by condition (17). The interquartile ranges for both Haz1 and Haz2 conditions were also identical. The Timer condition exhibited a much lower median value of 8 (because there were fewer steps in the transition of control) with an interquartile range of zero.

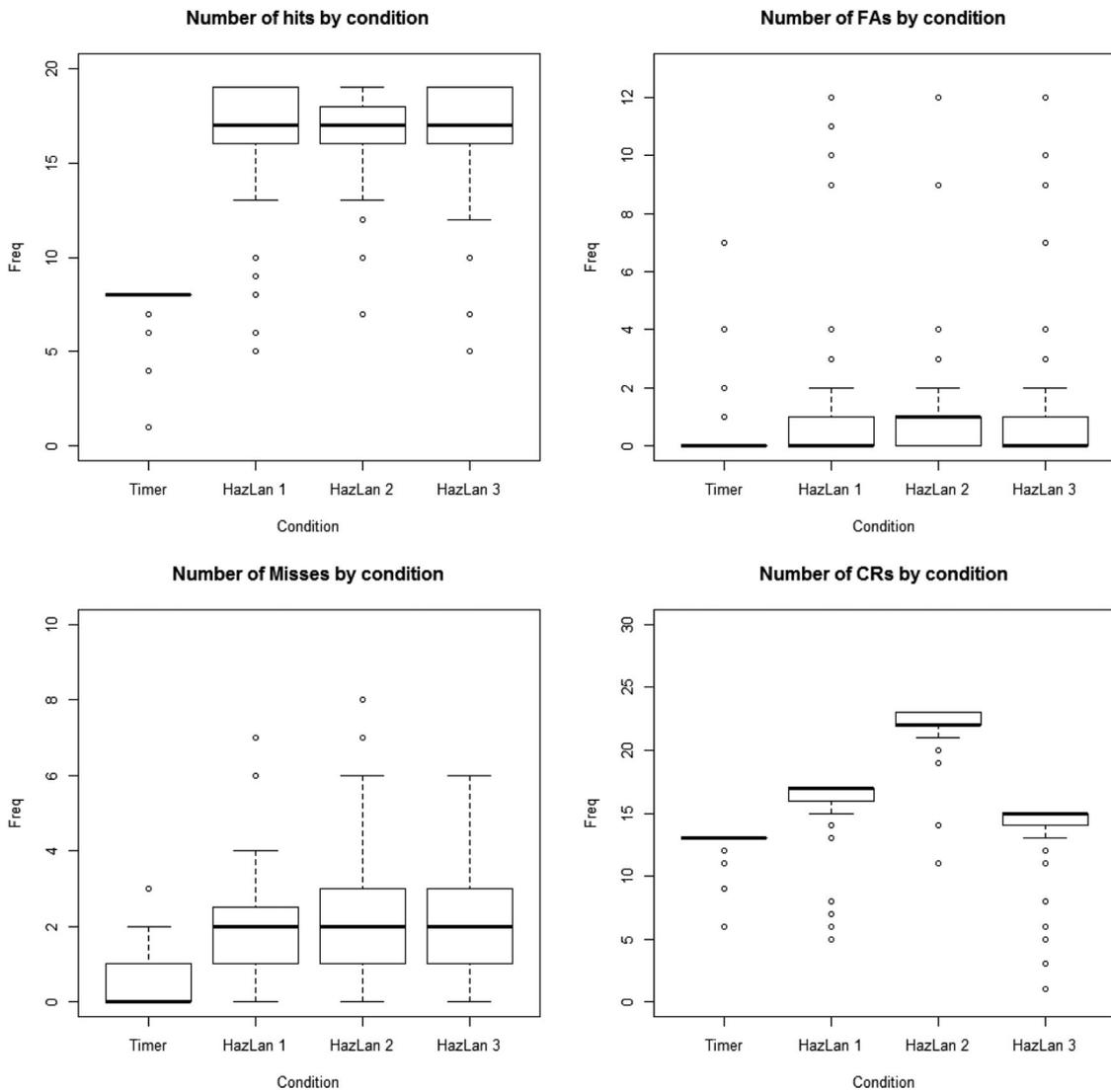


FIGURE 7 Higher-Fidelity experiment hits, misses, FAs, and CRs by condition. CR, correct rejection; FA, false alarm

Median values for misses were similar for all HazLan conditions at two and HazLan one and two conditions held the same interquartile values at two. The median value for the Timer condition was zero with an upper quartile range of one.

FA values by condition revealed similar median values of one for HazLan one and two conditions. HazLan three and Timer conditions both exhibited median values of zero. Interquartile ranges for all HazLan conditions were equal at one.

Number of CRs showed a variation between all four conditions. The Timer condition exhibited the lowest median, with a value of 13 and an interquartile range of zero. All HazLan conditions exhibited identical upper quartile ranges of one; the HazLan two condition had the highest median score (22) followed by the Hazlan one condition (16), and HazLan three (15).

Figure 8 illustrates a box plot displaying the Matthews correlation coefficient (ϕ) for each of the four conditions. Median ϕ levels above 0.8 suggest a strong positive relationship.

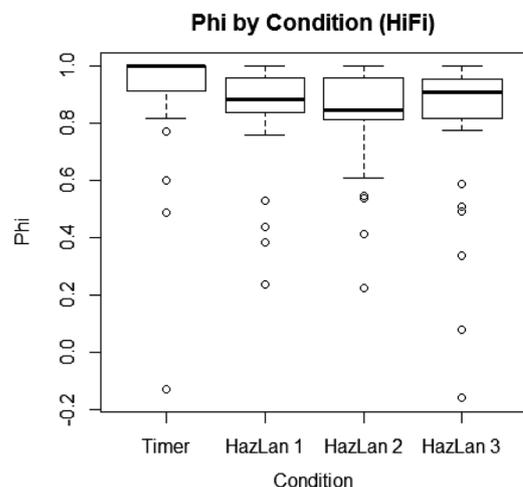


FIGURE 8 Phi (ϕ) by condition

5 | DISCUSSION

This paper sought to validate a method for modeling the handover from vehicle automation to the human driver. OESDs are offered as a way of predicting how drivers are most likely to behave during the process of handover of control in driving simulators. The two studies show that in both lower and higher fidelity driving simulators, the median ϕ statistic exceeded 0.8, which is the standard criterion for acceptable validation (Landis & Koch, 1977). It is interesting to note that the difference in the fidelity of the simulators did not affect the performance of the OESDs. This means that OESDs can be used to make predictions about vehicle–driver handover behaviors with some confidence. Over 100 drivers took part in the studies with an age range of 17–86. For both studies, there were a relatively low frequency of FAs (i.e., predicting behaviors that did not occur) and a high frequency of hits (i.e., predicting behaviors that did actually occur). This is coupled with a relatively low frequency of misses (i.e., failing to predict behaviors that actually occurred) and a high frequency of CRs (i.e., not predicting behaviors that do not occur). In addition, the reliability of the classification of behavior into the four categories of hits, misses, FAs, and CRs from the video data were acceptable for both studies. There is often a disappointing lack of evidence on the reliability and validity of human factors and ergonomics methods (Stanton, 2016), but this does not have to be the case. This and other studies have shown the way in which the evidence can be collected, analyzed, and presented (Stanton et al., 2009).

The findings from the current studies are promising for the continued use of OESDs, and it is certainly comparable with the better-performing methods in the discipline of ergonomics and human factors (Stanton et al., 2013, 2014). Other studies on the prediction of task performance time (Harvey & Stanton, 2013a; Stanton & Young, 1999a) and human error (Stanton & Baber, 2005; Stanton & Stevenage, 1998; Stanton et al., 2009) have all produced good levels of validity. This is not to say that all methods perform as well. In one of the first studies of its type, Stanton and Young (1999a, 1999b) compared a range of methods, and some performed quite poorly. Stanton (2016) notes that the better-performing methods are generally quite focused in terms of their predictions, such as time and error (and now activities). Nevertheless, it is important to understand the limitations of any method before using it.

The misses are probably the most interesting category of behaviors in this study, as they represent behaviors conducted by the driver in the handover of vehicle control that were not anticipated in the OESD modeling. These were unexpected behaviors that may have occurred due to the drivers being impatient or over-eager to take manual control of the vehicle. With hindsight, the early takeover behaviors could have been modeled in the OESDs. Vocalizing readiness in the phase before carrying out previous processes, such as ceasing using a secondary task or resuming driving position, was not modeled as it was assumed that the driver would only vocalize readiness when they were actually ready.

However, the incidence of this behavior suggests impatience with the duration of the verbal feedback interaction or lack of perceived value in completing the full procedure. Equally, behavior such as removing hands from the wheel during the protocol and having to replace them during or after the protocol was not modeled as this was not anticipated in the design of the interaction. It was assumed (wrongly in some cases) that once the driving position was adopted, it would be maintained. The duration of the handover protocol may have encouraged drivers to take their hands off the wheel and feet of the pedals (after first placing them there) as they did not need to manually steer until the handover procedure was complete. A design solution to overcome this could include the use of steering wheel sensors to detect the correct driving position and provide feedback if hands are removed after the commencement of the handover procedure. This demonstrates the utility of the OESD method to test design assumptions and highlight short-cuts or process failures as a means for generating mitigation strategies through design or training.

Although these studies focused on the use of OESDs to predict handover of vehicle control, the approach could be extended to other aspects of driver behavior (Banks et al., 2014). It has shown utility in other domains, such as aviation (Sorensen et al., 2011; Harris et al., 2015; Walker et al., 2010). As a cautionary note, however, validity generalization cannot be assumed and must be tested (Stanton, 2016). OESDs are good at modeling discrete events, such as the stages in the handover of vehicle control, but modeling continuous events is more challenging (such as manual control of a vehicle). This may require new notation and procedures for OESDs to handle the continuity of nondiscrete activities, such as steering, maintaining speed, maintaining lane position, searching for hazards, anticipating the behavior of other road users, reading the road ahead, and so on. The nomenclature would also require some notation for the interruption of the continuous activities (such as making an emergency stop) as well as some way of representing the duality of activity in driving (such as operating the climate system while also driving the vehicle). Harvey and Stanton (2013b) presented the multimodal critical path analysis method to show how the driver could engage in activities simultaneously (spearing cognitive and physical activities by modality). This would mean that OESDs need more than one “swim-lane” for the representation of driver behavior. Huddleston and Stanton (2016) borrowed notation from computer science to represent continuous activity. So, potentially at least, there are methods for extending OESDs to cope with both continuous and multiple driver activities. Future OESD modeling research could consider the incorporation of time and error data within the analysis (Stanton & Salmon, 2009) to predict VtD handover times as well as the exceptions. Data on human performance time are available in the literature and could be extended to this domain (Harvey & Stanton, 2013b; Stanton & Baber, 2008; Stanton et al., 2014). This has the advantage of helping vehicle designers budget time allocation for the VtD handovers, a topic of much debate (Eriksson & Stanton, 2017).

6 | CONCLUSIONS

OESDs appear to be able to make good predictions about driver activity during the simulated handover of vehicle control from automation in both lower and higher fidelity simulators. Over 100 drivers were tested in both studies with different interaction designs, and the median validity statistics were all above the criterion value. Consequently, the OESD method may be used with some confidence in modeling driver interaction for discrete events, although validity generalization requires testing. Further development work is needed to incorporate continuous activities into OESD modeling.

The debate on the reliability and validity of Human Factors and Ergonomics methods has been continuing for some time (Stanton & Young, 1999a, 2003). Very little work has been conducted to validate methods and, consequently, very little is reported in the literature (Stanton, 2016). Stanton (2016) presented a call to arms for the developers of Human Factors and Ergonomics methods to provide evidence that their methods really do all that is claimed of them, by way of formal, empirical studies of reliability and validity. The reporting that does occur in the open literature shows that there is scant evidence (Stanton et al., 2013). Examples include most of the methods used for task analysis, process charting, error and time prediction, team performance, and system design (Stanton et al., 2013). Much more could be done, and surely needs to be, for Human Factors and Ergonomics to be a credible engineering discipline.

OESDs could be applied in any domain where a description of the interactions between human operators and technology (including automation) would be useful. Examples of transportation domains include aviation, maritime, and rail. For example, the relationship between the autopilot and human pilots could be explored to improve the handovers and reduce the problems encountered in Air France 447 crash (Salmon et al., 2016). Another example could be using OESDs to explore the interactions between the signalman, the alarm system, and system control displays in the Ladbroke Grove rail collision (Stanton & Baber, 2008). In this latter case, the complexity of the interactions appeared to have slowed the reactions of the signalman. The OESD method is not restricted to transportation domains however and could be equally well be applied to control rooms in energy production and distribution, defense, manufacturing, medicine, nuclear, oil and gas, and security (Stanton et al., 2010). With the rise of interest in multiagent systems, OESD could prove useful in representing and exploring the relationships and interactions within and between different agents.

ACKNOWLEDGMENTS

This study was supported by Jaguar Land Rover and the UK-EPSCRC grant EP/N011899/1 as part of the jointly funded Towards Autonomy: Smart and Connected Control (TASCC) Program.

DATA AVAILABILITY STATEMENT

Due to the nature of this study, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

ORCID

Neville A. Stanton  <http://orcid.org/0000-0002-8562-3279>

REFERENCES

- Abdi, H. (2007). Signal detection theory (SDT). In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 886–889). Sage.
- Banks, V. A., & Stanton, N. A. (2017). *Automobile automation: Distributed cognition on the road*. CRC Press.
- Banks, V. A., Stanton, N. A., & Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind' free driving. *Safety Science*, 62, 505–514.
- Eriksson, A., & Stanton, N. A. (2017). Take-over time in highly automated vehicles: Transitions to and from manual control. *Human Factors*, 59(4), 689–705.
- Harris, D., Stanton, N. A., & 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, N. A. (2013a). Modelling the hare and the tortoise: Predicting the range of in-vehicle task times using critical path analysis. *Ergonomics*, 56(1), 16–33.
- Harvey, C., & Stanton, N. A. (2013b). *Usability evaluation for in-vehicle systems*. CRC Press.
- Huddleston, J. A., & Stanton, N. A. (2016). New graphical and text-based notations for representing task decomposition hierarchies: Towards improving the usability of an Ergonomics method. *Theoretical Issues in Ergonomics Science*, 17(5–6), 588–606.
- Hutchins, E. (1995a). *Cognition in the wild*. MIT Press.
- Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, 19, 265–288.
- Keates, S., & Clarkson, P. J. (2003, November). Countering design exclusion through inclusive design. In *ACM SIGCAPH Computers and the Physically Handicapped* (No. 73–74, pp. 69–76). ACM.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Moray, N., Groeger, J., & Stanton, N. A. (2016). Quantitative modelling in cognitive ergonomics: Predicting signals passed at danger. *Ergonomics*, 60(2), 206–220.
- Nevin, J. A. (1969). Signal detection theory and operant behavior: A review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. *Journal of the Experimental Analysis of Behavior*, 12(3), 475–480.
- Politis, I., Langdon, P., Adebayo, D., Bradley, M., Clarkson, P. J., Skrypchuk, L., Mouzakitis, A., Eriksson, A., Brown, J. W., Revell, K., & Stanton, N. A. (2018, March). An evaluation of inclusive dialogue-based interfaces for the takeover of control in autonomous cars. In *23rd International Conference on Intelligent User Interfaces* (pp. 601–606). ACM.
- SAE (2018) J3016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. https://www.sae.org/standards/content/j3016_201806/
- Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., 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, P. M., Walker, G. H., & Stanton, N. A. (2016). Pilot error versus sociotechnical systems failure: A distributed situation awareness analysis of Air France 447. *Theoretical Issues in Ergonomics Science*, 17(1), 64–79.
- Sorensen, L. J., Stanton, N. A., & Banks, A. P. (2011). Back to SA school: Contrasting three approaches to situation awareness in the cockpit. *Theoretical Issues in Ergonomics Science*, 12(6), 451–471.
- Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied Ergonomics*, 37(1), 55–79.

- Stanton, N. A. (2014). Commentary on the paper by Heimrich Kanis entitled 'Reliability and validity of findings in ergonomics research': Where is the methodology in ergonomics methods? *Theoretical Issues in Ergonomics Science*, 15(1), 55–61.
- Stanton, N. A. (2015). Responses to autonomous vehicles. *Ingenia*, 62, 9.
- Stanton, N. A. (2016). On the reliability and validity of, and training in, Ergonomics Methods: A challenge revisited. *Theoretical Issues in Ergonomics Science*, 17(4), 345–353.
- Stanton, N. A., & Baber, C. (2005). Validating task analysis for error identification: Reliability and validity of a human error prediction technique. *Ergonomics*, 48(9), 1097–1113.
- Stanton, N. A., & 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, N. A., Eriksson, A., Banks, V. A., & Hancock, P. (2020). Turing in the driver's seat: Can people distinguish between automated and manually driven vehicles? *Human Factors in the Manufacturing and Service Industries*, 30, 418–425.
- Stanton, N. A., Salmon, P., Harris, D., Marshall, A., Demagalski, J., Young, M. S., Waldmann, T., & Dekker, S. W. A. (2009). Predicting pilot error: Testing a new methodology and a multi-methods and analysts approach. *Applied Ergonomics*, 40(3), 464–471.
- Stanton, N. A., & Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, 47(2), 227–237.
- Stanton, N. A., Salmon, P. M., Jenkins, D. P., & Walker, G. H. (2010). *Human factors in the design and evaluation of central control room operations*. CRC Press.
- Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. (2013). *Human factors methods: A practical guide for engineering and design* (2nd ed.). Ashgate.
- Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: Situation awareness in individuals, teams and systems. *Ergonomics*, 60(4), 449–466.
- Stanton, N. A., & Stevenage, S. (1998). Learning to predict human error: Issues of reliability, validity and acceptability. *Ergonomics*, 41(11), 1737–1756.
- Stanton, N. A., & Walker, G. H. (2011). Exploring the psychological factors involved in the Ladbroke Grove rail accident. *Accident Analysis & Prevention*, 43(3), 1117–1127.
- Stanton, N. A., & Young, M. S. (1999a). What price ergonomics? *Nature*, 399, 197–198.
- Stanton, N. A., & Young, M. S. (1999b). *A guide to methodology in ergonomics: Designing for human use*. Taylor & Francis.
- Stanton, N. A., & Young, M. S. (2003). Giving ergonomics away? The application of ergonomics methods by novices. *Applied Ergonomics*, 34, 479–490.
- Stanton, N. A., Young, M. S., & Harvey, C. (2014). *Guide to methodology in ergonomics: Designing for human use* (2nd ed.). Taylor & Francis.
- Volvo Cars. (2015). Volvo cars: Explore the user-interface of tomorrow. [Video file]. <https://www.youtube.com/watch?v=xYqtu39d3CU>
- Walker, G. H., Stanton, N. A., Baber, C., Wells, L., Gibson, H., Salmon, P. M., & Jenkins, D. P. (2010). From ethnography to the EAST method: A tractable approach for representing distributed cognition in air traffic control. *Ergonomics*, 53(2), 184–197.
- Walker, G. H., Stanton, N. A., & Salmon, P. M. (2015). *Human factors in automotive engineering and technology*. Ashgate.

How to cite this article: Stanton, N. A., Brown, J., Revell, K. M. A., Langdon, P., Bradley, M., Politis, I., Skrypchuk, L., Thompson, S., & Mouzakitis, A. (2022). Validating Operator Event Sequence Diagrams: The case of an automated vehicle to human driver handovers. *Hum. Factors Man.* 32, 89–101. <https://doi.org/10.1002/hfm.20887>