Situation Awareness of Remote Vehicle Operators

Clare Mutzenich

Department of Psychology

Royal Holloway, University of London

Submitted for the Degree of Doctor of Philosophy, July 2022

Primary Supervisor:

Professor Polly Dalton (RHUL)

Secondary Supervisors:

Dr Shaun Helman (TRL)

Dr Szonya Durant (RHUL)

Declaration of authorship

Declaration of Authorship

I, Clare Mutzenich hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

Date:15/07/2022

Acknowledgements

When I applied for the PhD studentship in 2018, I wrote openly in my application that I aspired to play a part in creating psychological research, rather than teaching about it in a classroom. The last 4 years have made that desire a reality and I am humbled by the range of experiences, methods, and training that I have been able to access during the PhD to build my skillset. Thank you to SeNSS and ESRC for the generous funding, TRL for being such an involved industry partner, all the staff in the Psychology department at Royal Holloway and the PhD student body, past and present.

I want to give my heartfelt thanks to my supervision team who have supported me throughout.

Polly, from the very first phone call when I called to enquire about the studentship, all the way through to the final submission of the PhD project, you have inspired me, encouraged me, and supported all my goals (even the ill-fated MRI study!). To have such a wonderful supervisor is one thing, but I am incredibly fortunate to also be able to call you my friend, thank you for always giving me your time and energy. No one could have been more enthusiastic or motivating, especially when I had minor crises of self-belief. I know you will dedicate the same energy to helping me celebrate!

Szonya, thank you for all your help and support for the last 4 years and for encouraging me to do the eye tracking analyses at the end of the PhD. You made it sound so easy, I am glad that I had no idea what a challenge the data cleaning would be, but it was all worth it in the end! I had such fun working with you and can't wait to do more analysis together in the future.

Shaun, thank you for always asking me such hard questions! You made me think about what and why I was doing anything – I learned so much from your focus on the small details. Nothing was ever too much trouble for you to help me and you always made it so easy to ask for your assistance and advice. Thank you, again, for continually bigging me up to anyone who will listen!

I also want to say a massive thank you to Doga for your tireless focus and patience while we discussed the strategy and analysis for Paper 5, and for your manual efforts when we decamped with the whole VR lab to TRL's offices in December 2021.

Clare Lally, you deserve a special mention for helping me so much with R coding, particularly in my second year; I learned such a lot about how to approach coding problems from working with you. You really do write such cool code! Thank you for taking the time to go through my scripts and for always showing me how to improve them.

Simon, my incredible husband, you never stop telling me how successful I can be, no one has more faith in me than you. I am not saying it lightly when I claim I couldn't have done it without you – your collaboration has been emotional, practical, and often even technical! I could never have taken this opportunity without you supporting our family while I took a break from my career to achieve this PhD, yet you never had any doubt it would be worth it in the long run. You have truly seen the highs and lows during this whole process, and I trust and value your counsel more than anyone in the world. Thank you for belief and support, I love you.

This thesis is dedicated to Noah and Lorelei, my lovely children. I will give you some advice from what I have learned in the past 4 years - always listen most to the good things that people tell you about yourself; believe them and make them true. I love you both.

Abstract

In the future, as highly automated vehicles become more prevalent, a remote operator (RO) may be called upon if a problem arises that prevents a vehicle from navigating independently to its destination. ROs will inevitably have reduced situation awareness (SA) in comparison with an in-vehicle operator. This thesis examines how long is required for an RO to build up SA and whether SA can be improved by changes to the mode of presentation or the provision of certain types of additional information from the remote scene. Paper 1 presents the state of the art at the start of thesis with regards to the scope of an RO. Paper 2 adopts a qualitative methodology to develop a taxonomy capturing the key elements of people's reported SA for videos of driving situations. Paper 3 develops and validates a novel measure of SA in remote contexts which possesses four underlying constructs: spatial and environmental awareness, anticipatory hazards, dynamic driving actions and other road users. Paper 4 investigates whether the presence or absence of a rear-view feed and/or audio feedback is helpful during the construction of SA in remote contexts, finding worse SA performance in the presence of the rear-view feed. Paper 5 compares two formats of presentation of 360° videos (head mounted display (HMD) and screen-based) with eye tracking analysis. Performance on a choice reaction task is faster and more accurate when information is presented in HMD format and there are differences in eye movements in different parts of driving videos depending on presentation format. Paper 6 is an applied research collaboration examining the effect of type of screen presentation (flat monitor/curved monitor) on RO SA for forklift operation. Qualitative findings indicate that the potential to augment visual perception though variable video feeds is a strength in using remote operation in warehouse logistics.

List of contents

List of figures

Paper 1

Paper 2

Paper 3

Paper 4

Paper 5

Paper 6

List of tables

Paper 2

Paper 3

Paper 4

Paper 5

Paper 6

Introduction

1.0 Background to the research project

The introduction of automated vehicles (AVs) has, for several years, been hailed as an imminent revolution in the transport industry with the potential to change many people's daily lives. At the start of this PhD project in 2018, news reports confidently predicted that we would have driverless cars on UK roads by 2021, citing incentives ranging from reducing pressure on the existing highway infrastructure of towns and cities, supporting an aging population with innovative mobility solutions and limiting the number of road accidents that stem from errors in human judgement[1]. Both BMW and Ford confirmed plans to sell a fully autonomous vehicle with no steering wheel or pedals by 2021 and BMW capitalised heavily on ridesharing and ride-hailing services. This activity was echoed in the US, where multiple start-ups in California and Arizona, such as Waymo and Cruise advertised on-demand ride-sharing platforms and self-driving taxi services, albeit with a test driver ready to take over if required. This activity strongly suggested that over the next few years there would be an inexorable transition towards ever-increasing levels of vehicle autonomy.

At the time of writing, midway through 2022, although trials of autonomous mobility solutions are becoming increasingly common across a broad range of industries, such as in freight and logistics, agriculture, mining, and shipping, driverless cars on public roads have failed to materialise at the scale that was anticipated. In part, this can be attributed to the fact that the prevailing opinion has been that for a car to be fully automated, it must have no human involvement at all, yet the technology required for a fully self-driving car to be correct 100% of the time is still far in the future. One possibility to help these issues would be a remote operator (RO), able to provide assistance or drive the vehicle remotely if the automation fails.

A remote driver is defined by industry standards as a,

> "*driver who is not seated in a position to manually exercise in-vehicle braking, accelerating, steering, and transmission gear selection input devices (if any) but is able to operate the vehicle*" (SAE international, 2021).

There is evidence of increasing acceptance of the need for a remote operator within the transport industry as many automated vehicle and technology companies, such as Oxbotica, StreetDrone and Imperium Drive in the UK, and Zoox and Uber, have placed the role of the remote operator firmly as part of their business model for automated transport (Davies, 2019). There are many different use cases that apply to remote operations, for example fleet operation, passenger assistance or

---

[1] https://www.bbc.com/future/article/20211126-how-driverless-cars-will-change-our-world

assuming direct driving control, all of which will require the RO to first gain situation awareness (SA) of the location that the AV occupies.

There are many definitions of SA and the concept has been applied across countless domains, but In basic terms, SA is knowing what is going on so you can take action (Adam, 1993). The most cited definition is Endsley's 1988 model that defines SA as,

> *'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future'*
> (Endsley 1988a, b, p. 792).

Endsley divided SA into three separate levels of awareness, Perception, Comprehension and projection. Applied in driving contexts, perception can refer to awareness of nearby objects such as pedestrians, other vehicles or road signs, comprehension refers to understanding and interpreting the perceptual information, for example estimating the distance to nearby objects to calculate risk, and projection requires evaluation of the likelihood of future events, for example a collision with other objects or vehicles. However, in the face of rapid developments in what it means to be a 'driver' discussed above, there are likely to be significant differences between SA in normal driving contexts and SA in remote driving operations.

1.1 Rationale for the investigation

The transport industry urgently needs to consider how relevant information can be transmitted from the remote scene to ROs in order to enable them to build and maintain remote SA as effectively as possible, as well as investigating how much time it is likely to take to make decisions based on information from a remote scene. The scope of this research is to provide empirical evidence which will be useful as regulatory frameworks are established with regards to the training required for remote operation, the necessary equipment and technology, and a comprehensive inventory of the use cases under which we could expect remote operation to be carried out.

1.2 Research aims

When ROs are asked to provide assistance to an AV, they have to 're-join the loop', often doing so with zero prior awareness of the situation. This is further compounded by the second-hand information they are receiving from it and the degraded quality of information transmitted from the environment in comparison to in-situ driving. To fully engage with the task, it will be critical for ROs to develop a sense of presence in the remote scene, a subjective perception of being somewhere else (Georg, Putz, & Diermeyer, 2020). These issues all combine to result in a longer exposure time to the scene being necessary to build SA. These SA challenges are unique to operators of AVs.

The overarching aim of this PhD thesis is to investigate the information and timescale that is required for a remote operator to gain sufficient SA to remotely control, guide or supervise an AV. The primary objectives of this research project are to,

- Update our understanding of situation awareness in relation to remote operators of autonomous vehicles.
- Investigate how to effectively measure SA needs during remote operation.
- Contribute to an understanding of the optimal format information should be delivered to an RO to build and maintain SA quickly and safely in different vehicle remote operation scenarios.
- Investigate factors that will increase an RO’s sense of presence in the remote scene.

1.3 Progression of studies

Paper 1 provides a literature review detailing the state of the art in remote operation of automated road vehicles, together with a detailed explanation of the types of cases in which an RO may be expected to take over control of an AV, discussion of the roles of a remote operator and an exploration of the challenges unique to an RO in charge of a vehicle they are not physically occupying. It was published under the title, 'Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles' in 'Cognitive Research: Principles and Implications' journal in February 2021.

Given that remote driving SA is such a new field, to inspect the implicit and explicit mental processes that are underway while ROs build up a remote model of the environment, Paper 2 adopts a qualitative methodology, to uncover what people “see” in a remote scene when they are not constrained by rigid questioning, by asking participants to watch videos of driving scenes and describe out loud what is happening (to measure SA Comprehension) and what will happen next (to show SA Prediction). This is to enable understanding of the SA first principles that underpin SA in remote contexts. This manuscript was published in the journal 'Frontiers in Psychology' in November 2021 under the title ‘Situation Awareness in Remote Operators of Autonomous Vehicles: Developing a Taxonomy of Situation Awareness in Video-Relays of Driving Scenes’.

SA needs during remote operation will differ to SA requirements during in-situ driving, meaning that standard measurements to assess SA in driving may not be appropriate. How to effectively measure SA needs during remote operation represents an operational challenge in the autonomous transport field, which has been historically approached inconsistently. Paper 3 addresses the lack of standardisation in previous quantitative driving-related SA measures by developing and validating a novel measure of SA in remote contexts.

Remote operators will need to take control of the vehicles at short notice, without the chance for their SA to have built up over the preceding time and will have to build up a mental model of the remote environment facilitated by monitor view and video feed. A critical question is to uncover how information should be delivered to an RO in order to build and maintain SA quickly and safely which is addressed in Papers 4, 5 and 6. Paper 4 explores the influence on remote SA of two information sources presented in addition to the standard front camera feed, namely the presence or absence of a rear-view feed and the presence or absence of audio feedback. Paper 5 considers the effect of two formats of presentation (either head mounted display or screen-based monitor display) when carrying out remote supervision of automated vehicles using a choice reaction task and employed eye tracking analyses to uncover whether there are differences in distributions of fixations between different formats of presentation which correspond with improved SA performance.

Paper 6 is an applied research collaboration between Phantom Auto and Royal Holloway, University of London (RHUL) to identify factors contributing to operator performance, enabling product decisions that are based on empirical data with the aim of increasing remote operators' SA. Operators performed three loading and unloading tasks in a warehouse environment via remote operation, viewing the video feed presented on either a flat or curved monitor, with performance times compared across the two conditions. Paper 6 represented an opportunity to conduct real world testing of remote operator SA for forklift operation.

Method

A detailed explanation and justification of the methods used for each study is included in each of the papers with critical evaluation of their use in the study. The following sections provide an explanation of the progression from the initial stages of research development to the six final papers divided into qualitative and quantitative methodological approaches.

1.0 Qualitative approaches

1.1 Literature review

Paper 1 was written as a call to the field to acknowledge the significant differences between SA in normal driving contexts and SA in remote driving operations. It was published in 'Cognitive Research: Principles and Implications' journal in February 2021. An understanding of the challenges for ROs of AVs is an urgent research priority addressed by this PhD thesis.

Paper 1 set out the current state of the art in remote operation of highly automated vehicles in a comprehensive literature review; proposing revisions to current taxonomies which described the stages of automation to allow for the potential handover to an RO, outlining the roles and use cases of a remote operator and providing a detailed analysis of the challenges in SA for ROs with suggested approaches for improving remote SA. The broad scope of this review paper set the scene for many of the research questions addressed by the studies in this PhD thesis. For example, the detailed discussion of the types of edge cases that may cause problems for AVs and necessitate a handover to an RO in Paper 1 shaped the design and filming stages of the creation of the driving videos for Papers 2, 3 and 4 and the classification of driving videos into types of edge case in Paper 5 to assess their effect on reaction times and accuracy of decisions. Furthermore, Paper 1 outlined three separate roles and use cases of a remote operator which led to shaping some of the studies in the thesis to different use cases, for example the experimental task in Paper 5 was based on an RO providing remote assistance to an AV with no expectation that any remote driving would be required whereas Paper 4 investigated the type of information that could be provided to an RO which would be useful for all types of use cases. Finally, one of the suggested approaches for improving remote SA by using a virtual reality head mounted display in 360°, allowing the operator to control their field of view just by moving their head was tested in Paper 5 with a comparison between head mounted displays and screen-based presentation.

1.2 Verbal elicitation task

The qualitative methodology adopted in Paper 2 evolved after experiencing challenges trying to abstractly design quantitative questions which would dependably measure SA. Alternatively, a qualitative study developing understanding of 'what' people saw in a remote scene when they were not constrained by contrived questions would likely improve question design and had the potential to reveal differences between driving SA in remote contexts and SA gathered whilst in-situ. Paper 2 moved away from traditional "freeze and probe" quantitative techniques commonly used to measure SA such as the Situation Awareness Global Assessment Technique (SAGAT), instead using a qualitative verbal elicitation task to uncover what people see in a remote scene (Endsley, 1988). The verbal protocol used, which asks people to verbalise their thinking aloud whilst recording everything they say, has been shown to assist in identifying underlying concepts that are difficult to extract by quantitative methods (Jewitt, 2012). Transcribing the participant transcripts provided a rich and informative dataset of over eight thousand words.

1.3 Inductive thematic analysis

Paper 2 used Inductive thematic analysis (TA) to encode the qualitative data garnered by the verbal elicitation task. Inductive techniques involve searching across the whole raw data set to find "themes," which can be analysed to generate theories, rather than deductive approaches which search for evidence of pre-existing theories (Braun and Clarke, 2013). The inductive approach was chosen to limit the influence of a priori preconceptions of SA from the prevalence of Endsley's "three levels of SA" model which could have dominated the thesis direction (Endsley, 1988). Indeed, Paper 3 suggests that standard measurements to assess SA in driving may not be appropriate in remote operation. Instead, the inductive TA methodology used in Paper 2 generated deeper understanding of the semantic information that people can access in naturalistic driving scenes, which would have been missed by quantitative approaches, allowing the generation of a taxonomy of driving SA which was then used to create SA questions for the driving videos used in Papers 3, 4 and 5. A weakness of TA however, is that the researcher is required to exercise subjective judgements when determining what is a theme, for example how many occurrences have to be present to be determined a theme depends on an interpretation of whether the theme represents a critical point (Braun & Clarke, 2013). To balance against this threat, Paper 2 outlines all the coding decisions that were made whilst defining the themes of the taxonomy for transparency and the research team were all involved in stepwise classification of the themes and sub-themes to limit researcher subjectivity.

2.0 Quantitative approaches

2.1 Exploratory Factor Analysis

The aim of Paper 3 was to iteratively design, test and refine the SA questions to ensure that the new SA measure had construct validity before using it in further studies. An exploratory factor analysis (EFA) was used to confirm the relationships between the SA questions, which were based on the taxonomy developed in Paper 2, and to identify the total number of dimensions represented by the questions. EFA is widely used in the social sciences to validate new testing measures and is useful when developing new theories, such as SA in remote driving (Knekta, Runyon, & Eddy, 2019). However, there are no clear ‘rules’ in how to conduct factor analysis, and a researcher can argue there is a theoretical justification for most decisions which leaves this method open to subjectivity and bias (Ledesma et al., 2021). During the process of extracting factors, labelling them and testing for inclusion reliability, evidence-based guidance was followed and reported at each stage in Paper 3 to ensure there was a theoretical justification for all decisions made (Watkins,2018).

2.2 Eye tracking techniques

Paper 5 used eye tracking metrics as a further measure of SA as it has been shown that drivers’ visual attention to the roadway can indicate their engagement with the surrounding traffic and environment which will be important to measure in remote contexts transmitted via video feed (Merat, Jamson, Lai, Daly, & Carsten, 2014). Paper 5 investigated remote supervision using a choice reaction task to assess the effect of two formats of presentation, either head mounted display (HMD-360 condition) or screen-based monitor display (SB condition). Eye and head movements in the HMD-360 condition were compared to eye movements and mouse clicks in the SB condition to give further insight into how SA develops in remote driving contexts. Eye movement data in the HMD-360 condition was collected using Tobii Pro Lab software and an inbuilt infrared Tobii eye tracker in the HTC VIVE headset. Eye tracking in the SB condition was recorded using the Tobii Pro Nano screen-based eye-tracker. Fixation durations between HMD-360 and SB conditions in areas of interest (defined by dividing the 360° video image into four equal parts) in 60 driving videos were compared to determine the amount of similarity of participants’ search task strategy dependent on presentation format. However, there are noted differences in eye movement patterns when viewing passive information, such as watching videos, compared to actively engaging in driving tasks (Mackenzie & Harris, 2015). The task in Paper 5 recreates one RO use case where an operator has been asked to intervene by an AV in a supervisory capacity rather than directly taking driving control so any eye movement data collected will only be able to inform how decisions are made in that use case and not across the full functional scope of an RO.

3.0 Online studies

Papers 2, 3 and 4 used online testing techniques, using the Gorilla.sc platform which played the driving videos to participants and prompted them with SA questions after each video. Participants took part via a laptop or desktop computer and were advised that they must be in a quiet environment with no background noise or distractions. Online testing has become more conventional in recent years in Behavioural sciences as researchers can access large samples and simultaneously test subjects quickly which may not be possible under lab conditions (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018). In 2020, during the 'Covid-19 Pandemic', when research with human participants was halted at Royal Holloway, it was possible to conduct the study described in Papers 3 and 4 online, which meant that data collection could continue uninterrupted during this period. Furthermore, the complex, counterbalanced, randomized, presentation of the videos that was required by the design of these studies was readily facilitated by the experiment builder on the platform, limiting the chance of human error during the testing process.

4.0 Collection and analysis of behavioural data

Behavioural and perfromance data was collected in all papers using conventional methods in the field. Paper 2 collected ten participant recordings in response to two open questions, 1) "what is happening?" (SA comprehension) and 2) "what will happen next?" (SA prediction) for each of eight videos and qualitatively analysed them together as one set. Each transcript was divided into singular words and each word coded as an individual item. Similar words or concepts verbalized by participants were grouped together and recorded as themes or sub-themes to construct a taxonomy of the key elements of people's reported SA for videos of driving situations.

Papers 3 and 4 used the same data collection methodology, designing SA questions for sixteen original driving videos which measured a particular level of SA (perception, comprehension, and prediction). Accurate responses for perception, comprehension and prediction were summed to derive an overall SA performance including a novel analysis that incorporated confidence judgements into SA scoring (only counting a response as accurate if it was reported with high confidence on a Likert scale). We also collected data and conducted analyses on self-reported responses to presence and workload questions after each condition which were measured on a Likert scale. Paper 3 used exploratory factor anlaysis to validate whether the questions successfully measured remote SA and Paper 4 used a 2x2 repeated measures ANOVA inferential tests to analyse results.

Paper 5 used a choice decision task and measured performance by the time taken in seconds to press the corresponding button on a keypad indicating their decision. We also analysed the accuracy of their decision; the correct answer for each video was scored as 1 and 0 if incorrect. In Paper 6, operator performance was measured in seconds for the time it took operators to pick up a pallet to the time they put it down in the correct position. Qualitative data was also collected throughout the study using informal unstructured interviews and overt observation while operators carried out the task/s.

Paper 5 and Paper 6 use linear mixed effects models (LMM) to analyse the datasets collected during the study. This decision reflected, in part, the training progression throughout the PhD thesis, but a clear advantage of LMM analysis over ANOVA for these studies was that this type of analysis factors random effects of within participant designs, such as the effect of participant and the effect of the stimuli, into the model which partitions the variance associated with these differences explicitly. LMM can also account for missing data and unbalanced datasets which featured in both of these two papers, although there is contention with this type of analysis that deciding what is a fixed factor is widely open to interpretation (Magezi, 2015). The decisions that were made in Paper 5 and Paper 6 to determine the model specification is explicitly discussed in detail for transparency at every stage of the process. Finally, the number of observations and participants required for a well-powered experiment using LMM analysis was carefully aligned with the advised amount in the literature (Brysbaert, 2019).

5.0 Justification of mixed method approach

The process of moving between qualitative and quantitative methodologies in this thesis was essential to enable firstly an understanding of the principles underpinning driving SA, then to design novel measures to assess driving SA in remote contexts. The initial stages meant that the quantitative methodologies in the thesis were supported by theoretical evidence drawn from primary research specifically into driving SA. Applying the multidisciplinary approach that this thesis has adopted has provided insights into a relatively new field of research that would been impossible to collect using only experimental techniques.

Paper 1

Mutzenich, C, Helman, S, Durant, S & Dalton, P. (2021). Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles. Cognitive Research: Principles and Implications. 6, 9 (2021). https://doi.org/10.1186/s41235-021-00271-8.

Mutzenich *et al. Cogn. Research* (2021) 6:9
https://doi.org/10.1186/s41235-021-00271-8

Cognitive Research: Principles
and Implications

REVIEW ARTICLE **Open Access**

Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles

Clare Mutzenich[1*], Szonya Durant[1], Shaun Helman[2] and Polly Dalton[1]

Abstract

The introduction of autonomous vehicles (AVs) could prevent many accidents attributable to human driver error. However, even entirely driverless vehicles will sometimes require remote human intervention. Current taxonomies of automated driving do not acknowledge the possibility of remote control of AVs or the challenges that are unique to such a driver in charge of a vehicle that they are not physically occupying. Yet there are significant differences between situation awareness (SA) in normal driving contexts and SA in these remote driving operations. We argue that the established understanding of automated driving requires updating to include the context of remote operation that is likely to come in to play at higher levels of automation. It is imperative to integrate the role of the remote operator within industry standard taxonomies, so that regulatory frameworks can be established with regards to the training required for remote operation, the necessary equipment and technology, and a comprehensive inventory of the use cases under which we could expect remote operation to be carried out. We emphasise the importance of designing control interfaces in a way that will maximise remote operator (RO) SA and we identify some principles for designing systems aimed at increasing an RO's sense of embodiment in the AV that requires temporary control.

Significance statement

Personal motorised mobility is central to modern life. Autonomous vehicles (AVs) offer a range of potential benefits to society and to individuals such as mobility solutions for those who cannot drive themselves in the form of ride-sharing or autonomous taxi services, and reducing the number of road collisions that stem from errors in human judgement. AVs also provide plausible solutions to the issue of overcrowded highways as connected cars will communicate with each other and navigate an effective route based on real-time traffic information, making better use of road space by spreading demand (Department for Transport 2015). The 'Waymo Driver' self-driving taxi service is operating in California and has already accumulated over 20 million miles on open roads (Waymo 2020). GM owned AV 'Cruise' received a permit from the California DMV in October 2020 to remove the human backup driver from their self-driving cars and their 'Origin' prototype will have no steering wheel or pedals (California DMV 2020). This activity strongly suggests that the next few years will see a transition towards ever-increasing levels of vehicle autonomy. Yet the impression that driverless cars will mean there is no human involvement, since there is no human physically present in the vehicle, is a fundamental misconception (Cooke 2006). In reality, many problems can arise that would require a human operator to remotely assess and instrumentally correct or direct the automation as AVs are not able to perceive some information that humans take for granted (Adams 2007). An understanding of the challenges for remote operators of automated vehicles and considering them as a part of the automation process is thus an urgent research priority.

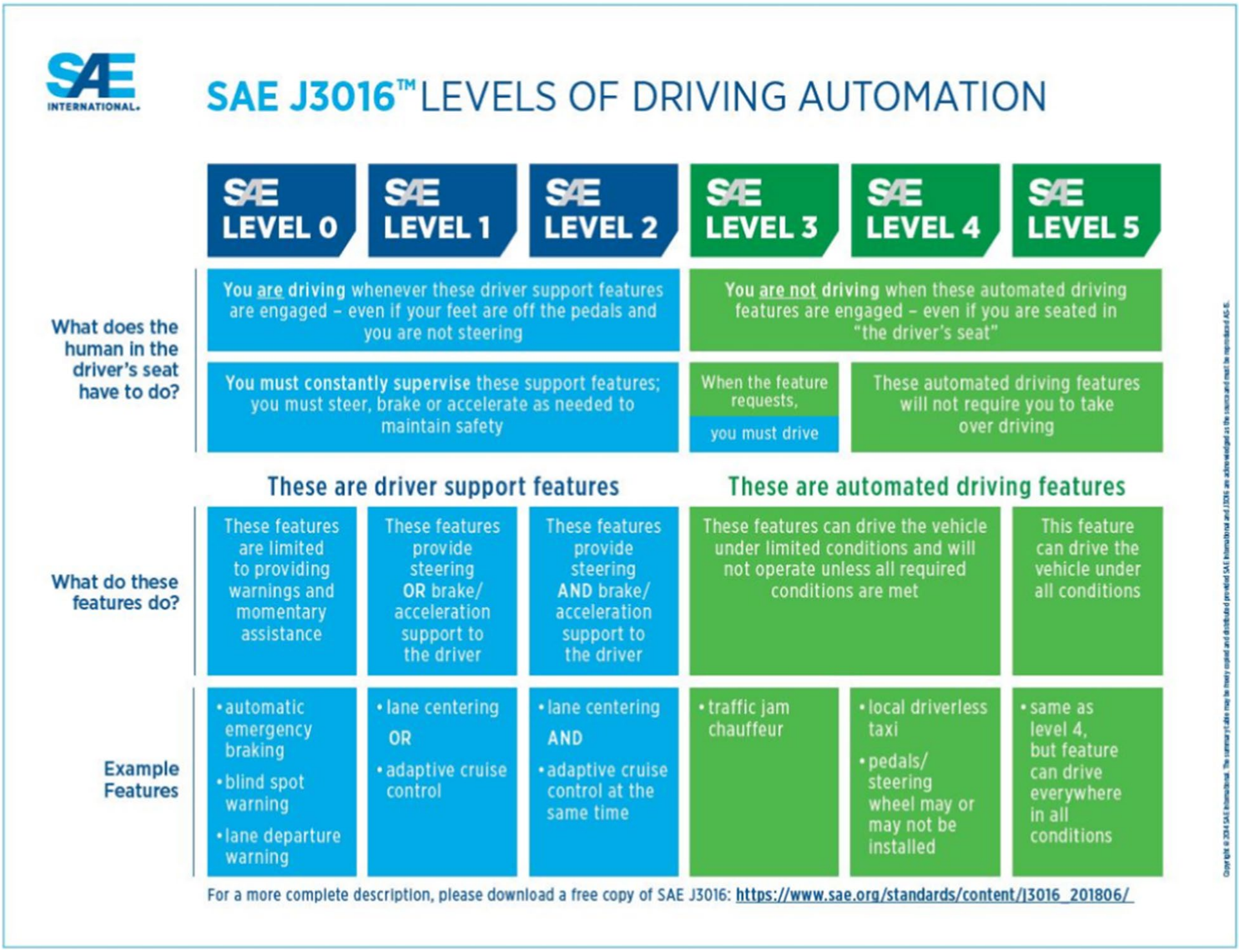
---

*Correspondence: clare.mutzenich.2018@rhul.ac.uk
[1] Royal Holloway, University of London, London, UK
Full list of author information is available at the end of the article

Introduction

Driving is an integral part of many people's lives—commuting to and from work, visiting friends or travelling across the country. The introduction of autonomous vehicles (AVs) could make more effective use of this time and prevent many accidents that are attributable to human driver error (Department for Transport 2014). However, even entirely driverless vehicles will sometimes require human intervention, which will often need to be provided remotely in the case of vehicles with no 'backup' driver present in the vehicle. This article, whilst not a formal systematic review, considers the extent to which our understanding of situation awareness requires updating to encompass these new contexts via a detailed examination of the current state of the art in remote operation of autonomous vehicles.

The organisational body SAE International (2016) highlighted six levels of automation for on-road vehicles (Fig. 1), with the aim of providing a universal taxonomy of terms for describing and defining levels of automation which can be adopted by industry, manufacturers and media. Levels 0–2 require the driver to be in charge at all times but with some partial assistance from enhanced or warning systems such as automatic braking systems. At Level 3 (conditional automation), the car can drive alone for short periods, merging onto motorways or changing lanes, however the driver is always physically present in the driver's seat, ready to take over if the car requests intervention. This assumes that the human is monitoring the driving environment either implicitly or explicitly and will be able to quickly re-engage with the driving process (Gugerty 1997, 2011). Tesla's model S offers a 'fully self-driving' mode which is, in fact, a Level 3 system as the driver is required to take over, meaning they must be in



SAE J3016™ LEVELS OF DRIVING AUTOMATION

	SAE LEVEL 0	SAE LEVEL 1	SAE LEVEL 2	SAE LEVEL 3	SAE LEVEL 4	SAE LEVEL 5
What does the human in the driver's seat have to do?	**You <u>are</u> driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			**You <u>are not</u> driving** when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			These are automated driving features		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering **OR** brake/acceleration support to the driver	These features provide steering **AND** brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering OR • adaptive cruise control	• lane centering AND • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

For a more complete description, please download a free copy of SAE J3016: https://www.sae.org/standards/content/J3016_201806/

Fig. 1 SAE International's (2016) six levels of automation for on-road vehicles, providing a universal taxonomy of terms for describing and defining levels of automation the driving seat and ready to respond in a timely fashion to takeover requests (despite high profile videos showing drivers sitting in the passenger or back seat (Mills 2018).

At Level 4 (high automation), the car can handle all dynamic driving tasks and should not require the human to take over driving. However at this level the AV is limited to its Operational Design Domain (ODD) and the car is programmed to achieve a minimal risk condition (MRC) by coming to a safe stop if there is a problem (Pollard 2018). The ODD may be controlled areas such as geofenced metropolitan zones or motorways, or may refer to a strictly defined route or be determined by weather conditions, speeds or time of day (Wood et al 2019). Level 4 AVs are likely to be autonomous shuttles which operate in small precincts or districts with a limited route and low speeds (less than 25 mph) such as the driverless shuttle trials offered by Oxbotica at Gatwick Airport (Oxbotica 2020).

By Level 5 (full automation), the passenger is required only to set the destination and start the car, as the automated driving system can operate full time performance of the driving brief. Levels 4 and 5 are differentiated by the fact that at Level 5 the vehicle is not restricted to an ODD and can operate on any road where a human could drive a car (SAE International 2018). The Automated and Electric Vehicles Act 2018 defines a vehicle as "driving itself" if it is operating in a mode where it is not controlled or monitored by a human (Law Commission, 2018, p. 9). Level 5 is seen as the final stage of automation where the car is fully self-driving and the human occupant would never be required to take over the driving.

Although there may be long periods of self-driving in a Level 4 or 5 AV, it seems disingenuous to expect zero system failure. This fact is widely recognised, with some industry experts, such as Waymo's CEO, even claiming that Level 5 autonomation is not achievable given the wide range of technical challenges involved in driving in "all conditions" (SAE International 2018, p. 2; Tibken 2018). The belief that humans can be finally "automated out of the loop" still proves to be a fundamental misconception, illustrating years of overconfidence in technology (Cooke 2006, p. 166). Problems will inevitably arise that are beyond the capability of the AVs' programming, obliging human involvement in the loop to assess the situation and instrumentally correct or direct the automation.

Until 2020, for AVs to be tested on public roads, legislation universally required that a safety driver must be inside the vehicle, at the wheel ready to take over if a disengagement was requested. However, changes to many European, US state and UK regulations have enabled a remote operator to assume this role. The Californian Department of Motor Vehicles (DMV) (the state that has the highest number of registered companies testing AVs on public highways) defines a remote driver as one not required to sit in the driver's seat. They may be able to monitor and communicate with the AV and may be able to perform driving or cause the AV to assume the MRC) which is usually coming to a safe stop (California DMV 2018). The United Nations Economic Commission for Europe (UNECE) considers remote operation as a key priority for regulation and has called for the definition of automated driving to be broadened to include remote support (UNECE 2020). In the UK in 2020, the Centre for Connected Vehicles (CCAV) regulated by the British Standards Institute (BSI), published two revisions to previous legislation, permitting remote operation to bring the AV to a controlled stop. The SAE Recommended Practice J3016 recognises and outlines the role of a remote driver as,

> *a driver who is not seated in a position to manually exercise in-vehicle braking, accelerating, steering, and transmission gear selection input devices (if any) but is able to operate the vehicle. (SAE International 2018, pg. 16).*

Furthermore, The Law Commission (2018) proposes that all AVs should have a person present who is able to take over the driving of the car if required, not a driver but a 'user in charge' who may be inside or outside the vehicle.

The handover from an AV to a safety trained human operator is referred to as a disengagement (Khattak et al. 2020). We can study data from disengagements to consider how frequently human operators may be required to re-join the loop. Currently, only the state of California records how many disengagements each company has per number of miles driven in that year and records the reasons for the disengagement, for example perception or planning discrepancy and further details such as weather conditions and location of disengagement. In 2019, Waymo, the self-driving AV of Google-owned company Alphabet, drove the highest number of miles (1.45 million miles) and recorded 110 disengagements (one per 13,182 miles). Sixty-one of these were related to AV perception issues for example, "failure to detect an object correctly", "incorrect behavior prediction of other traffic participants" and "adverse weather conditions" (California DMV 2019). Further examples of reasons for disengagements by all companies included poor traffic light perception (Phantom AI), sun glare (Mercedes-Benz), construction zones (Valeo), small debris in the road (WeRide.com) and "too big rain" (Nullmax). These types of programming deficits are known in the automation business as edge cases (Davies 2018).

Edge cases in autonomous vehicles

Edge cases vary significantly but they are colloquially defined as 'unknowable unknowns'—unpredictable, novel occurrences which fall outside the parameters defined by classifiers to recognise perceptual information (Koopman and Wagner 2017). They are triggered when an AV cannot correctly match the perceived information with known datasets, due to the presentation of ambiguous or incomplete stimuli. The neural networks that AVs rely on are trained on millions of images to enable the correct recognition and identification of a stimulus, however because edge cases are so unusual there are limited opportunities to train the system to recognise them (Hillman and Capaldi 2020). Gaps also exist in the datasets if an insufficient range of images have been used to train the algorithm, for example pedestrians may be labelled as walking on legs if disabled pedestrians were not shown in the training process, thus excluding wheelchair users (Koopman and Wagner 2016).

Edge cases can relate to animals, vehicles or objects presenting in unusual ways, for example a novelty trailer in the shape of a dog (Fig. 2) may be classified as an animal but its behaviour (travelling at speed on a motorway) may not correspond with system expectations which could trigger a minimal risk manoeuvre (MRM) such as an emergency stop. Further edge case examples may be generated by perception discrepancies caused by the target stimuli being occluded by other objects, such as a pedestrian moving behind a parked vehicle (California DMV Disengagement Reports 2019). Yet another example relates to the perception and interpretation of signage. The driving environment is frequently crowded with many signs cluttering both sides of the road offering competing information sources or even conflicting information. The Move_UK project found that some road signs only had a probability of detection of below 80%, mainly due to their location or angle, or whether they were obscured by vegetation or street furniture (Move_UK 2018). Human drivers are typically able to identify the relevant sign to their intended goal or destination and unconsciously filter out the other information, but an autonomous system may interpret all of them as relevant, particularly when it may not be apparent that they relate to a parallel road such as in Fig. 3. There is also evidence of consistent, systemic problems within many AV perception software systems to correctly 'see' or interpret cases of bare legs, children or red objects which raise significant safety concerns for pedestrians, vulnerable road users and traffic light adherence (Koopman 2020).

Fig. 2 An example of an edge case; a large model of a dog travelling on a car may be classified as an animal, but its behaviour (travelling at speed) may not correspond with system expectations. This Photo by Unknown Author is licensed under CC BY-SA

One approach to dealing with edge cases is to use human input in advance to teach an AV how a human would react in an emergency, but which can be applied by the AV in rapid time frames. The time it takes for an AV to assess the edge case could be reduced, meaning that time critical crashes could also be avoided, through the use of crowd sourcing, AI coordinated performance and reinforcement learning algorithms (Daw et al. 2019). The human is still required in this example to provide the human interpretation that the AV lacks, but they do so in advance, responding to randomly generated scenarios using simulation software which create a bank of potential actions that can be referenced by the AV. AVs

Fig. 3 An example of edge case related to multiple signs. An autonomous system may interpret all visible signs as relevant, even though some relate to a parallel road. "Furniture" by hartlandmartin is licensed with CC BY-NC-SA 2.0

predict risks frame by frame using vision-based software and if they see an unacceptable degree of perceptual uncertainty ahead they can access a library of suitable responses, eliminating the need to get in contact with an RO in that instance (Daw et al. 2019; Hampshire et al. 2020). It has been claimed that this type of assistance could have prevented the 2018 Uber crash in Arizona in which the system took 6 s to recognise the pedestrian pushing a bike as it did not conform to its expectations of cyclists or pedestrians, only determining 1.3 s before the collision that emergency braking was required (Daw et al. 2019; NTSB 2018). Even though this may forestall some edge cases relating to identification of stimuli, there are still many occasions where a human would need to remotely intervene to interpret the situation and determine the best course of action.

The likelihood of an AV encountering one of these edge case events grows with each passing mile, with the 676 autonomous vehicles registered in California alone in 2019 driving a total of 2,880,612 miles in 2019 (California DMV 2019). Even if an edge case was only encountered 0.01% of the time, that still represents a potential episode every 288 miles, although the technical software capabilities of AV companies vary significantly. Growing numbers of businesses offer services to stress test AV systems to find the edge cases in their software by assessing the "what-ifs", weaknesses and false negatives in the system, so that AV designers can mitigate those risks (www.edge-case-research.com). Developers can also use simulation software to focus on difficult driving scenarios and reproduce edge case scenarios, however these are still limited to human imagination and accordingly, in the words of Elon Musk, Tesla CEO, "simulation….does not capture the long tail of weird things that happen in the real world" (Wevolver 2020, pg. 40). Governments in the UK and the US have provided funding to research realistic edge case scenarios in simulated and real-world data. For example the D-risk project, part of a consortium of five organisations led by US start up, aiPod Limited, was awarded a £3 m grant in 2018 to use AI to develop a novel scenario generator by combining actual edge case scenarios (Centre for Connected & Autonomous Vehicles 2018). All this action illustrates the seriousness with which the industry is considering the impact of edge cases on Level 4 and 5 AVs.

Despite the continuing efforts to improve perception software in AVs, it is still likely that even if the AV is programmed to assume the MRC within its ODD and come to a safe stop, the car may still represent a risk to other road users or will need to be delivered to its destination. Indeed, there are still arguments as to whether the ODD at Level 5 is truly unlimited, as it may be unable to handle extreme weather conditions if its sensors fail (SAE J3016, 2016). Although a human may also struggle in some edge case scenarios, we possess the higher-level skills to interpret and react to novel scenarios. A scenario which represents an edge case for an AV may be easily interpretable by a human driver, suggesting that current automation technology still necessitates collaboration between humans and AVs (Hancock et al. 2019).

In a Level 3 (conditional automation) AV it is possible to take control by grabbing the wheel or hitting the brake. However, the future design of many Level 4 and all Level 5 vehicles could possibly have no steering control or brake at all. For example, the U.S. National Highway Traffic Safety Administration has recently approved the design and use of cars without steering controls and GM Cruise's new model, 'Origin', has no cockpit at all (Hawkins 2020; NHTSA 2016). In these types of AV, a human occupant would be unable to operate the car even if they wished to do so and so may need to call upon the services of a remote operator. Furthermore, there is disagreement amongst industry professionals as to whether the failsafe of performing the MRC in a Level 4 AV if the occupant cannot take over is an appropriate policy if the AV comes to halt in a line of traffic (Thorn et al. 2018). This may not in all cases represent a good strategy for the vehicle, its passengers or other road users, obliging some type of intervention by a remote operator who may be able to move the car to a more suitable location.

The functional scope that a remote operator offers can span from merely advising the AV of how to proceed, to interaction with passengers or to the extent of taking over the driving of the AV from line of sight or a remote location. The next section comprises an examination of the use cases under which we could expect a remote operator (RO) to be utilised and the roles or tasks the position may entail.

Roles and use cases of a remote operator

The three roles that a remote operator may be called upon to provide for an AV in the event of an edge case can be seen in Fig. 4. A remote operator (RO) may be required to provide **remote assistance** to a Level 4 and Level 5 AV by alerting the service provider when it has broken down or providing information and customer service to passengers (UNECE, 2020). For example, if the vehicle has a flat tire, a breakdown vehicle may need to be called and updates communicated to passenger as to how long the repair will take, whether an alternative vehicle will be dispatched etc. This type of service is already offered by GM's OnStar Crisis Assist program and AV companies such as AnyConnect and Sensible4 currently deploy remote control centres that are equipped to respond to customer service requests such as if a passenger demands to speak with an operator (Cummings et al.

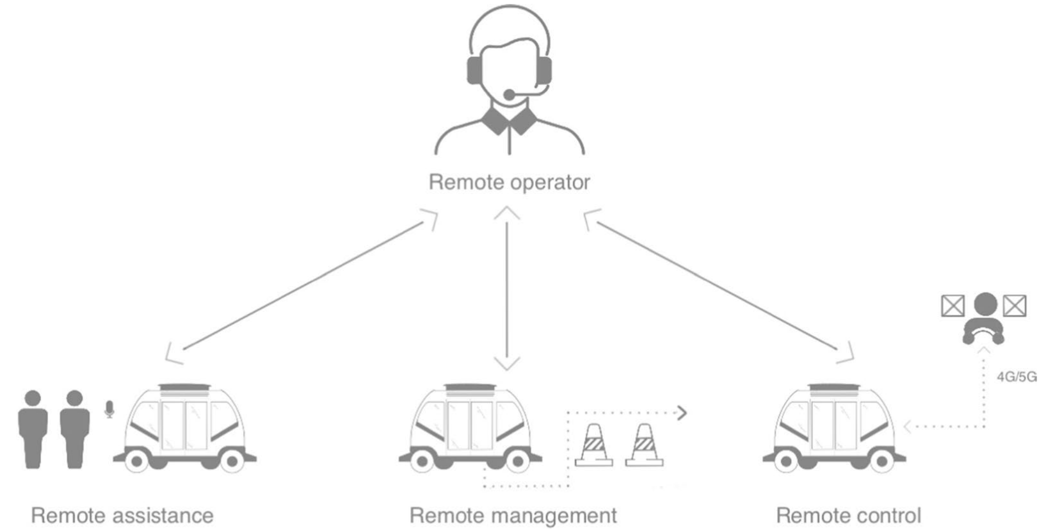


Fig. 4 Visualisation of the roles a remote operator may provide to an AV in the event of an edge case

2020, Sensible4.fi, AnyConnect.com). An RO may also be required to be responsible for the safety of passengers in self driving ride share situations, where there may be no conductor and the behaviour of other passengers may be a personal security risk (although this creates problematic surveillance and privacy issues for passengers on board which may need to be addressed by an on-call button rather than continuous audio monitoring). Even the simple task of opening and closing doors could be carried out by an RO or responding to a passenger request for an emergency stop (UNECE 2020).

A further role that could be offered by a RO is that of **remote management**, similar to an air traffic controller, where an RO working in a remote management post could also assume control of a fleet of AVs, as it would be poor economics to operate on a 1:1 basis (Hampshire et al. 2020). Fleet operations could include dispatch services which coordinate deliveries, navigational support and monitoring of radio and traffic information, for example communicating recent road closures to the entire fleet, as connected cars may wrongly interpret the low traffic volume as indicating the fastest route (Cummings et al. 2020). Giving the remote controller governance to order the AV to move or deviate from a fixed path would also enable the AV to override highway rules in exceptional circumstances, for example if instructed to do so by a police officer (UNECE 2020). Allowing driverless cars to call upon a centralised remote call centre means that several cars a day could be assisted using a blend of human experience and machine execution (Daw et al. 2019).

Furthermore, an RO may only need to assist in a purely advisory capacity in the event of an edge case triggering the MRC in a Level 4 or 5 AV (UNECE 2020). The RO could review the reason for the MRC and, after assessing the environment using the AV's cameras and sensors may confirm it is safe to proceed. Zoox is currently adopting this remote management strategy using 24 h support centres in which ROs provide human guidance to the AV, still in autonomous mode, in response to ambiguous scenarios (Zoox.com). The Texas A&M University shuttle program has introduced a self-driving trolley in the downtown precinct of Bryan, Texas, which is remotely monitored from a call centre where a RO authorises a restart when the shuttles is forced to stop for a pedestrian or object in the road (Costlow 2019).

This type of goal-based supervision could also be delivered in the form of a set of instructions e.g. for the AV to remove itself from the line of traffic where it may have assumed a MRC. It represents a prompt resolution to a vehicle obstructing traffic, but only requires basic training and carries less risk of human error than directly assuming control (Cummings et al. 2020). Nissan has also approached the challenge posed by edge cases by integrating remote management into its AV business model with its strategy of Seamless Autonomous Mobility. Operators working in a call centre, who Nissan refers to as Mobility Managers, plot a path for the AV around the object using a drawing interface and then return control to the AV to execute the path (Daw et al. 2019). The AV then relays the information to all the connected cars in its system, so each AV has that template to refer to in future similar situations, eventually reducing the need for an RO.

Lastly, the full test of an RO's capabilities would be assuming sole responsibility for the dynamic driving task either at low or high speeds from a remote location separate to the physical environment where the AV is located (UNECE 2020). This type of **remote control** is referred to as teleoperation and could be a viable solution to expand the ODD of Level 4 and 5 AVs (Thorn et al. 2018). Teleoperation is a substitute or enhancement to driverless transport, using 4G or 5G mobile networks to continuously video stream visual data from cameras around the car and linking driving operations to a driving rig in a remote control centre via an on board unit in the car (T Systems 2020). The remote 'driver' receives real time information displayed on multiple monitors or through a VR headset and can control all acceleration, deceleration, steering and braking using a traditional steering wheel and pedals, or joystick (UNECE 2020).

Many companies are already offering teleoperation as a mobility solution for autonomous services: Ottopia supplies teleoperation capability for both indirect control (such as remote management already discussed) and direct control of AVs, partnering with fleet operation provider, Bestmile; Phantom Auto have employed their teleoperation software to remotely control unmanned vehicles, delivery bots and forklift trucks since 2017 in Mountain View, California; and WeRide in China have removed the safety driver and are insistent that remote driving is the next step in making AVs profitable (Dai 2019). Additionally, there were an increasing number of start-ups registered in 2019 that included remote teleoperation in their business model such as Scotty Labs, who partnered with Voyage supplying self-driving cars in retirement communities (Dai 2019). Six states in the US expressly allow for teleoperation and England, Canada, Japan, Finland and the Netherlands have also authorised its use in supporting autonomous vehicles.

However, there is intense debate within the automation industry as to what extent teleoperation as a service is a viable option, with companies such as TuSimple and 2GetThere rejecting it as an inherently unsafe prospect, others, such as Nissan, who consider an RO as a necessary precaution to edge cases, but do not include direct control, or others like Zoox and Einride who are factoring in remote operation of the AV but currently only in some instances/locations. There are also key differences between current teleoperation practices as to whether remote driving is delivered only at low speeds (less than 25mph) such as EasyMile electric shuttles or at high driving speeds such as Designated Driver who successfully remotely operated a car at Goodwood Racecourse (Cummings et al. 2020; Designated Driver 2019). It seems probable though, in the future, that some or all forms of remote operator will become an important feature to support autonomous driving. Thus, we need to reflect on the safety measures, performance requirements and the key issues that will be relevant to operators of remote vehicles.

The current article addresses two main issues with regard to the remote operation of an AV. Firstly, SAE International's (2016) taxonomy does not acknowledge the possibility of remote handovers so suggestions are made to update the nomenclature. Secondly, ROs will face significant challenges that are unique to their role as driver in charge of a vehicle that they are not physically occupying. ROs are likely to require longer exposure times to gain sufficient situation awareness; they face latency and perception issues and may have difficulty achieving a sense of embodiment in the remote vehicle. We reflect on these issues and offer practical design features which may enable the RO role to be carried out within a safe and realistic framework.

An extended taxonomy of automated driving systems

As described earlier, the SAE International levels of automation are adopted by all manufacturers to synchronise the classification of AVs and describe their capabilities at each level (SAE International 2018). We argue, however, that the SAE taxonomy does not reflect the fact that ROs will occasionally be obliged to intervene in the driving of an AV, for instance in the event of an edge case as previously discussed. The expectation has always been that the SAE taxonomy will change as the industry itself evolves, so we submit that the taxonomy now needs to be extended to include remote intervention by a remote operator who is effectively part of the AV system.

We propose a revision to the SAE taxonomy to allow for the potential handover to an RO by the addition of Levels 3b, 4b and 5b to the existing Levels 3, 4 and 5 (see Fig. 5).

These adjunct levels encompass the three roles of the RO that we have outlined above (i.e. assistance, management and control), which are labelled collectively as 'remote intervention'. The levels of automation have been informally reduced to '*feet off*' [Levels 0–2] as the car can take control of the pedals, *hands off* [Level 3] as the driver does not have to touch the steering controls, *eyes off* [Level 4] for when the driver no longer has to watch the road and '*brain off*' [Level 5] as the occupant could even fall asleep as the AV would never require them to take over (Kemp 2018, p. 7). The extra level that we propose could be summarised as '*far off*', as the RO is intervening from a separate location possibly hundreds of miles away.

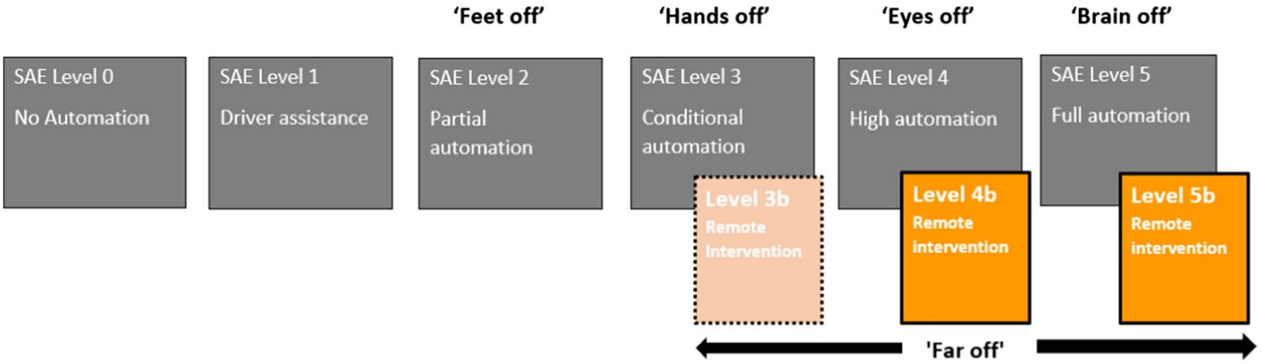


Fig. 5 Suggested revisions to the SAE (2016) taxonomy to allow for the potential handover to an RO by the addition of Levels 3b, 4b and 5b to the existing Levels 3, 4 and 5. The extra level proposed could be summarised as 'far off', as the RO is intervening from a separate location (nb. 'feet off, hands off, eyes off, brain off' summary taken from Kemp 2018, p. 7)

Remote intervention could potentially occur in the future even at Level 3 (indicated by the dotted lines and paler shading in Fig. 5) which is sometimes referred to as the 'mushy middle' of automation due to the AV's capability to drive itself but with an occasional need for human assistance. Furthermore, at Levels 4 and 5, where there is no human 'fail safe' within the vehicle, edge cases should be expected to be more frequent, indicated in Fig. 5 by the solid line for Levels 4b and 5b.

These additions represent a more comprehensive transition between system and remote human than the current SAE taxonomy for Levels 4 to 5 reflects, as it implies that the system **alone** is carrying out the execution of steering and acceleration/deceleration, monitoring of the driving environment and fallback performance of the dynamic driving task as is shown in Fig. 6. Instead, at Levels 4 and 5, the taxonomy should acknowledge the possibility of a reciprocal handover between system and remote operator, thus providing official recognition that edge cases will necessitate human–robot interaction for some time to come.

It is imperative to integrate the role of the remote operator within industry standard taxonomies so that regulatory frameworks can be established with regards to the training required for remote operation, the necessary equipment and technology, and a comprehensive inventory of the use cases under which we could expect remote driving to be carried out. An understanding of the unique challenges that remote operators of autonomous vehicles will encounter is subsequently an urgent research priority. We discuss these issues in the next section.

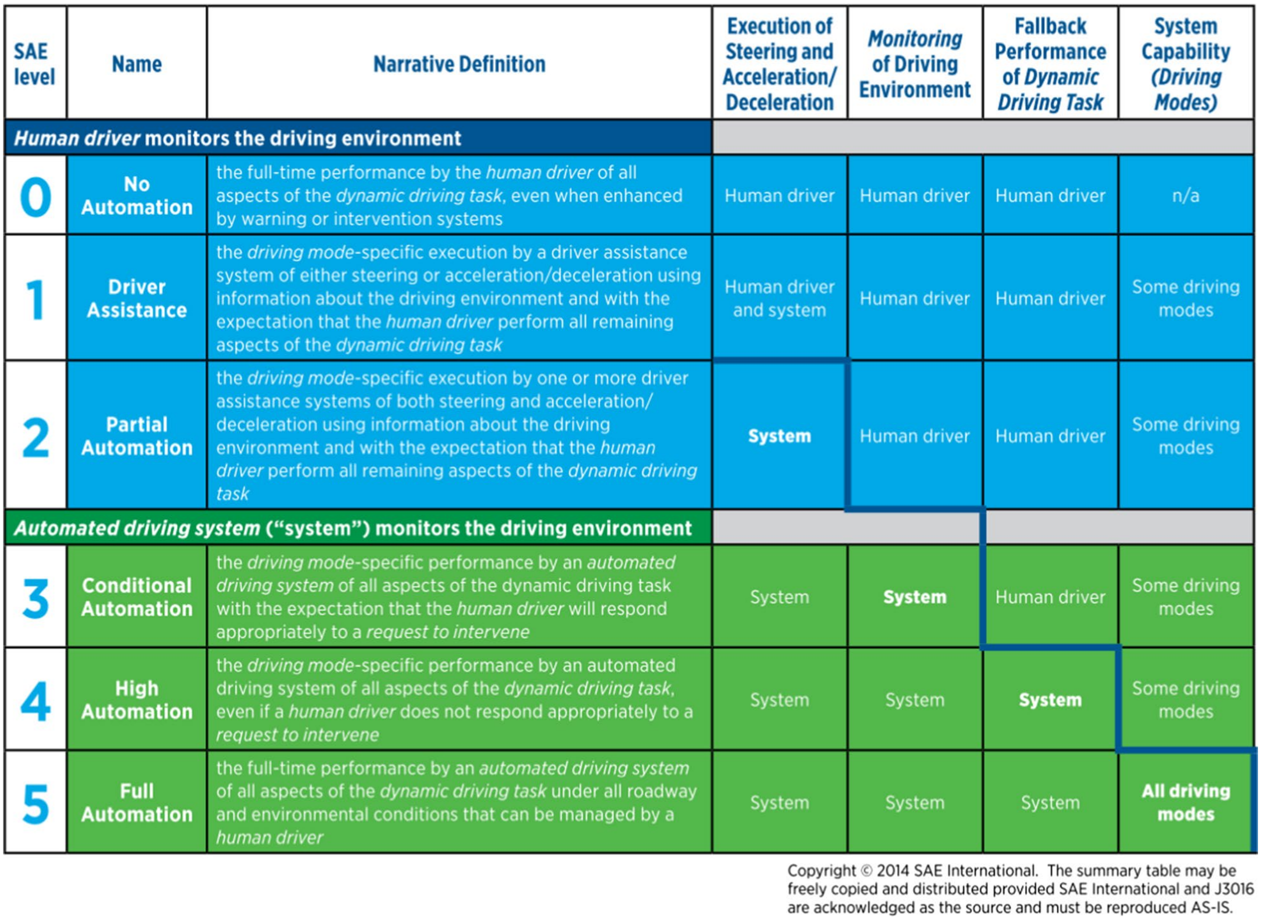
Situation awareness in driving contexts

There are many definitions of SA (see Endsley et al. 2003; Gugerty 1997, 2011; Lo et al 2016; Niklasson et al. 2007, Endsley 2015) but the most commonly cited is from Endsley's original model;

> *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future (Endsley 1988a, b, p. 792).*

Put more simply, SA fills the gap between what is known about the environment, what is happening in it and what might change. Endsley further divided the mechanisms of SA into three levels of responsiveness; Level 1, 'Perception' is the basic level of awareness that makes up the recognition of cues in the environment. Level 2, 'Comprehension', requires the current situation to be analysed, taking into account multiple pieces of information and their relative value to each other to make sense of what we are seeing. Level 3, 'Projection', a serial product of Levels 1 and 2, is the ability of the operator to make predictions about the future status of objects in their environment.

Endsley's SA model is well-established in numerous domains and has proved to be applicable to driving contexts (Bolstad et al. 2010; Endsley 2019; Ma and Kaber 2005). Indeed, inadequate SA is frequently implicated in crashes; failed to look/distraction, a Level 1 SA error, is the most common citation in insurance documentation (Department for Transport 2014). The SA perception



SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* (“system”) monitors the driving environment**						
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	**High Automation**	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Fig. 6 SAE International’s (2014) summary of the responsibilities of human driver and system at each level of automation for the execution of steering and acceleration/deceleration, monitoring of the driving environment and fallback performance of the dynamic driving task

requirements for Level 1 for driving would include being aware of the *presence and location* of nearby objects such as pedestrians, other vehicles and objects and/or road signs, what time of day or night it is, and current weather conditions that may be hazardous (Endsley 2020). Awareness of the *distance* to nearby objects, vehicles in the blind spot and the traffic lane needed takes the SA state up to Level 2 (Comprehension) together with the *impact* of the weather and road conditions on vehicle safety (Endsley 2020). Perception and comprehension awareness are constantly updated during the driving process as the environment is dynamic and both the actions of the driver and other road users will affect the ongoing analysis of the situation. Driving also necessitates SA Level 3 projection of the likelihood of *collision* with other objects or vehicles, and estimated *times and distances* to turns or exits (Endsley 2020).

The three levels of SA can be mapped on to how an AV views and interprets the driving environment. For Level 1, in AVs, the automated system ‘perceives’ via sensors such as LIDAR, RADAR and multiple cameras which can see ‘through’ walls and under the surface of the road, although limited to visual and auditory inputs (Kemp 2018). However, the AV sensors can be fooled, and false positives can lead to emergency braking manoeuvres. For example, in the first phase of the Move-UK project, the AV mistook a cloud of exhaust smoke hovering over the street as a solid object and instructed the vehicle to stop, showing that human intervention may be necessary even at the level of simple perceptual judgements as they may require a degree of synchronised comprehension (Seidl 2018).

In terms of Level 2, ‘Comprehension’, AVs do not currently possess the level of artificial intelligence necessary to achieve the nuanced comprehension of humans. Many edge cases are context dependent and an AV may fail to detect details that a human would know are either important or irrelevant (Koopman and Wagner 2016). For example, a human driver may edge forward slowly through a crowd of pedestrians blocking the road to exert their right of way but would know that this behaviour was not appropriate if the crowd was surrounding a casualty on the road.

This contextual distinction may not be as plain to a self-driving car so the unusual crowd behaviour would trigger a disengagement due to "interference of autonomous driving path by a pedestrian" (California DMV 2019; Law Commission 2018). This is when an RO could be required to take over from the AV in line with the remote management role outlined earlier to fill the "awareness gap" and either allow the AV to take a different path circumventing the crowd or instruct it to continue with the MRC (Drury et al. 2006, p. 91).

For Level 3, 'Projection', AVs currently struggle to make predictions with certainty as real-world driving is unpredictable and requires proactive as well as reactive decisions to avoid hazardous situations from developing (Endsley 2020). Humans are also unable to see into the future with certainty but we are capable of some anticipation, and of acting quickly and imaginatively in response to even unanticipated events; until AV software can demonstrate projection abilities at the same or at a greater level to those of a human driver, the requirement for remote operation is likely to remain, since an edge case is likely to still occur.

BSI guidelines in the UK, mentioned earlier, stipulate that the RO must be as safe, with the same level of situation awareness and response time, as a human safety driver assuming manual control of the car in the ODD (BSI 2020a, b). Research has attempted to quantify how much time it takes for drivers to build up SA in non-automated driving. Humans are capable of visually processing a natural scene within milliseconds as we can quickly pick up the gist of the contents (Thorpe et al. 1996). Lu et al. (2017) played participants videos of varying lengths and asked them to reproduce the traffic layout they had seen of the three-lane road. They found between 7 and 20 s was necessary to build up the necessary SA to complete this perceptual task successfully. However, when required to assess the relative speeds of other cars in relation to the ego vehicle, participants took 20 s or more. Fisher et al. (2007) found that the time it takes for drivers to become aware of latent or impending hazards in a scene was around 8 s and, in simulated scenarios, participants take around 12 s to feel safe enough to take over manual control of driving (Coster 2015). This suggests that perception of the world around you occurs quickly (Level 1 SA) but an understanding of what others are doing in that environment is slower (Level 2 SA) (Endsley 2017a, b).

The need for a new understanding of remote SA

An RO who has been alerted by an AV to drop in and assess an edge case or assume direct driving control will first need to acquire SA of the remote scene, yet there are significant variations between SA in normal driving contexts as we have outlined above and SA as it might occur in remote driving operations. The task of developing SA from a remote location is likely to be made more difficult by the operator's physical absence, however ROs will also have access to additional information (for example, from the advanced sensors on the vehicle in question, and from the sharing of information across entire fleets) such that some aspects of their SA will be enhanced by comparison to traditional driving. We accept Endsley's Levels 1/2/3 as the core basis of how SA is constructed but argue in this article that a new consideration is needed in order to encompass the scenario of a remote 'drop in' to an AV.

Endsley's model is most often considered as the operator, be it a military pilot or a road vehicle driver, being in-situ and experiencing information first-hand. In contrast, an RO is likely to suffer from a degraded state of SA as they are being transmitted indirect cues unlikely to replicate the full range of visual, vestibular, auditory and temporal information that is available to a driver in situ. Endsley clarifies her position by stating that SA is "the internal model of the world around [...] at any point in time" (Endsley 1988a, b, p. 789). However, this cannot logically apply to an RO, as the world around them will be very different to that which they are experiencing though the video feed of the AV. For example, a forward camera view has reduced motion parallax information, which, together with a reduction in image quality, will reduce the depth cues available to a remote viewer, by comparison with someone situated within the scene itself. Similarly, the audio relayed to the RO from the scene (if any) will be of reduced quality and presented against the background noise of wherever the RO is physically present. However, on the other hand, the scope of cameras and other sensory information provided by an AV may give an RO superior awareness of some aspects of the environment, highlighting information that would be easy to neglect in person or is beyond the visual capability of humans.

Furthermore, Endsley (1988b, pg. 3) describes SA as "an *ongoing* process achieved by accumulating knowledge over the course of the mission" combined with the feedback from the *immediate* environment [our italics]. This will not be the case for remote operation of Level 5 cars as it is not feasible or efficient to monitor all cars on all roads constantly on a 1:1 basis. Instead, the most likely scenario would be a centralised control hub that AVs can contact when they encounter an edge case, which function using systems analogous to air traffic controllers; supplying remote management but also potentially delivering real-time remote operation and even 'look ahead' models which draw on human and artificial intelligence simulated interactions to cache pre-determined responses to potential situations (Daw et al. 2019). This type of teleoperation safety service can assist several cars a day and is already being offered in different forms by Silicon Valley start-ups, such as Phantom Auto, and autonomous trucking companies across Europe, such as Einride (Davies 2018).

Accordingly, although an RO will undoubtedly need to build up SA, their SA is not *ongoing*. An RO will 'drop in' to a scene having no prior exposure, meaning that they will need to develop SA from scratch, without access to previously accumulated knowledge for a particular vehicle and context. Neither does the RO occupy the 'immediate' environment that the car does. The likelihood of ROs having to unexpectedly take control of AVs in a wide variety of unfamiliar locations makes it essential to identify how their SA needs will be different from those of an on-site driver.

SA demons for ROs

The difficulties of building and maintaining SA, often referred to as 'SA demons' have been widely documented in driving contexts, such as attention tunnelling, cognitive overload and out of the loop (OOTL) syndrome together with anxiety, fatigue and perception errors such as change blindness and errant mental models (Endsley 2012). However, there will be challenges in relation to OOTL syndrome, latency, embodiment, and workload that are specific to ROs of highly automated vehicles. We discuss each in turn in the next section and make suggestions as to how these SA risks can be mitigated.

Out of the loop syndrome (OOTL)

SA is likely to develop differently for an RO compared with that of a driver who is present within the vehicle as they will be 'out-of-the-loop'. OOTL is a major consequence of automation, leaving operators of automated systems handicapped in their ability to take over manual operations quickly in the event of automation failure (Endsley and Kiris 1995; Jones and Endsley 1996; Ottesen 2014; Porathe et al. 2014; Radlmayr et al. 2014). An RO in a remote call centre will be OOTL and it will take precious time for them to determine the cause of an edge case and successfully intervene (Endsley 2020). Attaining good SA in unknown remote environments is likely to be challenging for operators and they will experience a potentially hazardous delay while they build SA. Forming SA is also more challenging under time pressure, yet studies have found that operators will suspend all other tasks for up to 30% of their time to gain and re-gain sufficient SA, showing that maintaining SA is almost as difficult as attaining it (Drury et al., 2003; Larochelle et al. 2011, Porathe et al. 2014; Yanco and Drury 2004). An RO cannot begin to take control until they have built up adequate SA. This has serious implications if the location in which the AV has assumed the MRC presents a hazard for other road users, yet how much SA is 'enough' to start driving is hard to define even for a driver inside the vehicle.

Previous research into OOTL problems in highly automated driving have focused on the time taken to build up SA after take over requests (TORs) in Level 3 vehicles (for example Gold et al 2013, 2016; Lorenz et al. 2014; Melcher et al 2015; Radlmayr et al 2014; Salmon et al. 2012). Mok et al. (2017) found that 5–8 s are necessary to take back control of a Level 3 AV, after being engaged in an active secondary task (playing on a tablet). Eriksson and Stanton (2017) found that response times on average for on-road driving take over from Level 3 automation were around 3 s. However, as participants in Level 3 AVs are still inside the vehicle, sitting at the wheel, even with the distraction of the task it can be assumed that they still had some implicit SA which would not be available to an RO in a separate location (Gugerty 1997). An RO will be both cognitively and visually 'blind' prior to the TOR, so it is reasonable to assume there will be a longer delay for them to build up SA.

Likewise, in instances where the disengagement request from the AV is well-defined, for example 'sensor failure', ROs will be able to respond more quickly than if the AV cannot provide a cause, for example 'perception error' (California DMV 2019). In line with this assumption, Scholtz et al. (2004) found in field trials of semi-autonomous vehicles in multiple terrains that it took around 29 s for the remote operator to gain SA when it was specified that they needed to assist because of 'Plan Failure' (for example a plotted course did not succeed because the terrain was too bumpy). Yet, on average, it took 162 s to build up SA when the robot requested operator assistance but was not able to specify the cause of the problem. We can gauge from this that even when the RO knows why they have been asked to intervene, the time it takes to build up the necessary SA to take action is not trivial, but that far longer may be required in the event of an edge case where the RO has to work out the cause of the TOR.

Latency issues for ROs

All autonomous driving is made possible through the use of mobile phone networks which transmit data (T Systems 2020). Assuming remote control over a self-driving car requires high amounts of data transfer and broad network coverage together with low latency (i.e. as small a delay as possible in the time it takes for a signal to be transmitted, in the case of an AV and teleoperated driving, from the car to an operator situated miles away). For an RO to be able to drive the vehicle in real time the time-lag between the signal and response of the car must be minimal otherwise turning, accelerating, and braking will all be delayed. Any latency of over 50 ms will mean that the image the RO is seeing is out of date in terms of useful interaction (T Systems 2020). This has led to debate in the industry as to whether remote management or teleoperated driving is even a genuine possibility as any latency will have detrimental safety effects on the operator's ability to build time-critical SA.

For remote operation to be viable, AVs must be able to guarantee consistent, real time streaming of all relevant data to the control centre to enable ROs to build and maintain SA. Proponents of teleoperated driving argue that there are already test cases that confirm latency is not an issue for RO SA on 4G networks. For example, Scania controlled a bus remotely using multiple camera feeds which reduce the need for high Megabits per second (Mbit/s), and Designated Driver teleoperated a car on the English coast from their offices in Portland, US (Ericsson 2017). However, the road environment frequently encounters obstacles to good network coverage such as tunnels, overhead trees and 9% of the UK, mainly in rural areas, are unable to get 4G coverage. This is a significant problem for remotely controlling AVs that require at least 50 Mbit/s to stream visual data from four HD cameras around the AV (OFCOM 2019).

Over 80 towns and cities across the UK now have 5G capability which has the potential to reduce latency to less than 10 ms (from 50 ms on 4G). Connected and autonomous vehicles can use 5G wireless services to send information between themselves and to remote call centres at faster speeds and with higher capacity, supporting RO's ability to quickly gain SA. 5G will also enable priority service provisioning to vehicles currently being teleoperated which will reduce the risk of network dropout and promises to pave the way for teleoperated driving, at any speed, to become part of AV business models for telecoms companies when adopted nationwide (Ericsson 2017; T Systems 2020). However, no matter how fast the transmission between AV and RO, restricting the sensory information that the RO is receiving to solely visual, the main type of information we have considered so far, will have an impact on their sense of embodiment in the vehicle. It may be critical to provide additional modes of sensory information to enhance the RO's immersion in the remote scene when building SA.

Embodiment issues and SA for ROs

Myriad human–robot disciplines have identified potential SA problems relating to missing sensory information (Ottesen 2014). Remote ship operators provide a good example in this respect. In rough seas, the autopilot is often disengaged and the ship steered by hand, enabling the handler to 'feel' the ship's movement; this is not possible for the remote driver (Porathe et al. 2014, Jones and Endsley 1996). Similarly, when we manually drive a car, we feel 'part' of the vehicle even ducking instinctively as we go under a low bridge or 'sucking in' as we squeeze through a narrow space. An RO, not being physically present, is likely to miss this sense of *embodiment*; they cannot feel the seat beneath them or the pull of the wheel in their hands, they are in no personal danger and they are likely, until 5G is nationwide, to receive all visual and auditory information with at least some level of time-lag.

Even more concerning, limitations on ROs' SA stemming from a lack of embodiment may result in a sense of detachment or reduced perception of risk (UNECE 2020). Remote operators have cited a sense of driving 'deaf' or feeling like it is a game, when the reality is that they are potentially driving real passengers with the resulting consequences if they crash only borne by those at the scene (Davies 2019). Even ROs offering remote assistance to passengers, without undertaking any remote driving, may lack the empathy or rapport that an on-board safety driver or conductor may share with fellow passengers (UNECE 2020). Although they may have access to a wider range of sensors from the AV system than if they were manually driving the car at the location, they have no vestibular feedback and so may misunderstand the conditions 'outside' or attribute greater significance to one piece of information than another (Endsley 2019). A lack of embodiment may also have a deleterious effect on speed perception; without force feedback pushing you back into the seat or information from the tyre friction on the road, it is difficult to accurately judge how fast you are driving and to remain engaged in the driving task (Tang et al. 2013). This may be addressed to some extent by remote operation training which teaches operators how to pick up cues from other feedback, for example spatial audio from microphones placed around the car, multiple viewpoints available at the same time, above, around and behind the car and enhancing video feeds with camera 'blur' to simulate speed cues or pop up with speed warnings (Störmer 2019; Tang et al. 2013; UNECE 2020).

Workload issues for ROs

Unfortunately, one potential downside of providing this type of additional data is that the task of remote operation may then begin to push the boundaries of the "SA demon" of cognitive overload (Drury et al. 2007; Ottesen 2014, p. 2). Each additional piece of information provided carries a processing burden for ROs, who need to absorb the information and decide how to act. However, it is also theoretically possible that problems could arise due to a workload that is too low. For example, a situation in which the RO only has to deal with a low number of disengagements during a shift may exacerbate OOTL issues such as decreased vigilance. A careful consideration of the ways in which workload interacts with SA will therefore be essential in ensuring the safety of remote intervention. This would ideally disentangle the different types of workload which are assessed by standard measures such as the NASA Task Load Index (NASA-TLX), for example the mental, physical and temporal demands placed on the operator by the task and their resultant effects on performance, effort and frustration (Hart 2006).

Even the ways in which operators are allocated to jobs, including prioritising new AV requests when an operator is in the middle of a current call, will have important consequences for the operators' workload (Daw et al. 2019). RO roles such as remote management, where ROs are in supervisory control, may allow easier division of attention across multiple assignments than direct teleoperation which requires total focus on a single vehicle (Cummings et al. 2020). Indeed, some remote management requests may be experienced simultaneously by many cars, for example in the event of an accident on the road, in which case they can be dealt with together with the same instruction, reducing workload. However, the workload could easily become too high under these conditions if too many vehicles are allocated, making the risk of errors more likely (UNECE 2020). Waymo is tackling this challenge by allocating separate support teams to different operations such as fleet technicians, fleet dispatch, fleet response and rider support which share the workload and allow for job specialisation (Waymo 2020).

Different forms of autonomy will also require more or less intervention than others and will therefore impose different levels of workload on ROs. Consider, for example, the difference between long distance autonomous trucking and local delivery robots. Although the robots operate at low speeds, perhaps implying a lower workload than high speed truck driving, teleoperating the 'last mile' local environment (as opposed to long distances on motorways) is more likely to involve busy or crowded situations, such as in a loading dock or fuel station, which creates higher demand on SA (UNECE 2020, StarskyRobotics 2020).

Carrying out remote operation work is likely to be stressful and highly specialised, thus there is an urgent need for training and regulation of ROs to ensure safe performance in this challenging and demanding role (Hampshire et al. 2020). Exploring technological developments which can make the interfaces used by ROs more intuitive, for example VR and head mounted displays (HMD) can reduce workload and improve RO SA (Hosseini and Lienkamp 2016). We discuss potential solutions to the challenges we have raised for operator SA in the final section.

Suggested approaches for improving remote SA

In remote environments, it may be beneficial to make an RO's driving experience as realistic as possible whereby visual, auditory and haptic cues are provided as if they were in-situ. An operator's sense of embodiment relates both to their sense of location, feeling that they are occupying the same environment as the AV, and to their sense of body presence, feeling that they are physically inside the vehicle through sensory experience (Pamungkas and Ward 2014).

A naturalistic experience of driving even in remote contexts could be supplied by using a virtual display headset, allowing the operator to control their field of view just by moving their head (Almeida et al. 2014). This would avoid the need for multiple 2D monitors showing different camera feeds, which are likely to increase workload demand (Ricaud et al. 2017). Virtual Reality (VR) creates an illusion that the user is viewing the remote environment from an egocentric perspective, which seems likely to improve an ROs sense of embodiment in the scene. Their SA can also be enhanced by VR to provide a RO with a 360 view of the surrounding environment by combining LIDAR data with visual information from cameras and AR presented in a headset (Hosseini and Lienkamp 2016) Indicative of the potential success of applying VR to remote driving, Hyundai has released a prototype for a remote control system that allows an operator to drive a vehicle using binocular camera feeds to a VR headset that would give the vehicle operator a 3D view of the car's immediate surroundings (United Lex 2017). If ROs are able to combine teleoperation with a sense of telepresence by using VR technology this should logically decrease the time it takes to build SA, by effectively narrowing the time–space detachment that has previously limited operators (Almeida et al. 2014). However, there are motion sickness issues with VR which develop as the vestibular input does not match the visual motion experience of the user (Hosseini & Lienkamp 2016). Current technological developments under design to address motion sickness include presenting a view of a fake steering wheel or the operator wearing VR gloves which show the driver's hands in the VR environment which aligns visual and motion cues (Manus 2020).

Augmented reality (AR) may facilitate better SA in ROs as it can provide extra information superimposed over the visual information provided via cameras to the RO, for example showing possible navigational paths that the operator could take (Brandão 2017). If the AV has had a failure on a busy motorway, AR could help the RO navigate the AV through three lanes of traffic safely by highlighting a safe route. Enhancing the video stream using AR can also improve operators' depth perception and reduce workload as all the information is available in close proximity (Ruano et al. 2017). Placing AR virtual markers overlaid onto a map would allow ROs to 'see' salient information coming up which may be occluded by forward terrain or buildings (Ruano et al. 2017). NASA has tested a display system for cockpits that uses GPS to map an accurate picture of ground terrain creating an electronic picture that is unaffected by weather or time (NASA 2017). This would be invaluable when the remote environment is unfamiliar to the RO and the 'drop in' is unexpected and ambiguous, possibly under poor weather conditions. An RO, because of their remote location to the AV, would struggle to anticipate visual elements that are not directly apparent in the limited video feed. Augmented reality (AR) has been used to facilitate better SA by creating a "synthetic view" superimposed in the view of the RO using a heads-up display unit showing possible navigational paths that the operator could take that may be blocked from their direct view (Brandão 2017, p. 298).

Potentially, an RO of an AV that has unexpectedly requested assistance could request a short video summary that would show what happened seconds before to assist SA comprehension of the scene. 'Evocative summaries' have been applied in multiple domains to present key visual information quickly and may provide helpful insights as to how best to present information to an RO (Russell and Dieberger 2002). Recent AV software company, AnyConnect, can give operators access to 4G/5G video, audio, and data recordings from a few minutes before the event, although it would depend on different use cases whether it would be more or less valuable to an RO to have a snapshot of what happened in comparison to images from the current scene.

Equally, ROs can gather more data from the environment around the remote AV if they have control over the cameras, which may be necessary in edge cases that are related to ambiguous information in the roadway. Giving ROs the autonomy to assess only parts of the environment that are of interest to them may reduce the exposure time it takes to build remote SA. However, the benefits of operator control are debatable, as humans are prone to errors. In the DARPA AAAI-2002 Robot Rescue Competition, one team's operator, during the first run of the arena, moved the robot's camera off-centre to look at something more carefully. However, he then forgot to tap the screen to reposition the camera back to the centre, which meant he thought that the orientation of the camera was looking forward when it was actually 90° to the left. This SA error resulted in him driving the robot into the crowd (Yanco and Drury 2004).

Poor spatial and navigational awareness will reduce SA so providing ROs with pre-loaded terrain data with the AV's current position superimposed on it, will give ROs better comprehension of 3D spatial relationships. However, driving is an inherently visuo-motor task so the video feed cannot be replaced with map tracking information solely. Although map-centric interfaces have been proven to be more effective in providing good location awareness, video-centric interfaces appear to improve awareness of surroundings so a combination of both on a graphical user interface would be ideal (Drury 2006). Many cars already have navigation systems as standard that show where the car has come from and their planned destination, so this information could be effectively transmitted to ROs to enable them to build stronger SA of the AVs location, spatially and navigationally (Yanco and Drury 2004). However, HD maps carry worrying safety failure rates as they may not update planned or unplanned road changes or unauthorised changes to road signs (Wood et al 2019).

It is important to reiterate that many of these suggestions for improving SA in ROs involve presenting additional information. Their benefits must therefore be carefully weighed against the additional workload that they will impose. We need to determine which is the most relevant information needed by ROs and how this can be transmitted to them efficiently, allowing them to build up SA quickly and effectively without risking cognitive overload. Future empirical research will be essential in exploring these issues to develop user interfaces that pull together concepts from research in other, related areas and applies them to the specific case of ROs of AVs. This work is currently underway in our laboratory.

Abbreviations

AV: Autonomous vehicle(s); SA: Situation awareness; RO: Remote operator; MRC: Minimal risk condition; ODD: Operational design domain; BSI: British Standards Institute; VR: Virtual reality; AR: Augmented reality.

Acknowledgements

The authors would like to thank Nick Doyle Design for the kind contribution of Figure 4.

Authors' contributions

CM wrote the manuscript. PD, SD and SH reviewed several drafts and contributed to revisions of the manuscript. All authors approved the final manuscript.

Funding

Funding source—ESRC Doctoral Training Partnership with South East Network for Social Sciences (SeNSS). Grant Reference Number: ES/P00072X/1.

Availability of data and materials

Not applicable to this review article.

Ethics approval and consent to participate

Not applicable to this review article.

Consent for publication

Not applicable to this review article.

Competing interests

There are no competing interests for any of the authors.

Author details

[1] Royal Holloway, University of London, London, UK. [2] Transport Research Laboratory, Crowthorne, UK.

Received: 6 August 2020 Accepted: 9 January 2021

References

Adams, J. A. (2007). *Unmanned vehicle situation awareness: A path forward* (615). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.4867&rep=rep1&type=pdf.

Almeida, L., Patrao, B., Menezes, P., & Dias, J. (2014). Be the robot: Human embodiment in tele-operation driving tasks. In *IEEE RO-MAN 2014—23rd IEEE international symposium on robot and human interactive communication: human-robot co-existence: adaptive interfaces and systems for daily life, therapy, assistance and socially engaging interactions* (pp. 477–482). https://doi.org/10.1109/ROMAN.2014.6926298.

Bolstad, C., Cuevas, H., Wang-Costello, J., Endsley, M. R., & Angell, L. S. (2010). Measurement of situation awareness for automobile technologies of the future. *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety, 4970,* 195–213. https://doi.org/10.4271/R-402.

Brandão, W. L. (2017). Using augmented reality to improve dismounted operators' situation awareness. *IEEE Virtual Reality (VR), 2017,* 297–298.

BSI. (2020a). *Assuring the safety of automated vehicle trials and testing-Specifi cation Publishing and copyright information*. Retrieved from https://www.bsigroup.com/globalassets/documents/pas/pas1881_final-design-proof.pdf.

BSI. (2020b). Guidelines for developing and assessing control systems for automated vehicles. *BSI Standards*, *PAS 1880:2*.

California DMV. (2019). *Disengagement Reports 2019*. Retrieved from https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/.

California DMV. (2020). DMV authorizes cruise to test driverless vehicles in San Francisco. Retrieved November 5, 2020, from https://www.dmv.ca.gov/portal/news-and-media/dmv-authorizes-cruise-to-test-driverless-vehicles-in-san-francisco/.

Centre for Connected & Autonomous Vehicles. (2018). *UK Connected & Autonomous Vehicle Research & Development Projects*. Retrieved from ccav.org.uk website: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737778/ccav-research-and-development-projects.pdf

Cooke, N. (2006). Human factors of remotely operated vehicles. In *Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 166–169).

Coster, X. A. (2015). *A method for determining how much time people need to understand traffic situations* (Delft University of Technology). Retrieved from https://repository.tudelft.nl/islandora/object/uuid:4225c90b-9911-4c50-b32c-892315e4c440.

Costlow, T. (2019). Far and away: Remote drivers monitor autonomous vehicles. Retrieved from https://www.sae.org/news/2019/10/remote-drivers-monitor-autonomous-cars.

Cummings, M. L., Li, S., Seth, D., & Seong, M. (2020). *Concepts of operations for autonomous vehicle dispatch operations*. 32. Retrieved from https://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u36/CONOPSdispatch_Final_compressed.pdf.

Dai, S. (2019). Remote control technology could be the key to moving autonomous driving forward, WeRide CEO says. *South China Morning Post*. Retrieved from https://www.scmp.com/tech/enterprises/article/3031585/why-remote-control-technology-could-be-key-moving-autonomous.

Davies, A. (2018). Self-driving cars have a secret weapon : Remote control. Retrieved October 19, 2018, from https://www.wired.com/story/phantom-teleops/.

Davies, A. (2019). The war to remotely control self-driving cars heats up. Retrieved October 13, 2020, from Wired website: https://www.wired.com/story/designated-driver-teleoperations-self-driving-cars/.

Daw, A., Hampshire, R. C., & Pender, J. (2019). *Beyond safety drivers: Staffing a teleoperations system for autonomous vehicles* (pp. 1–55). Retrieved from http://arxiv.org/abs/1907.12650.

Department for Transport. (2014). Contributory factors to reported road accidents 2014. *Reported Road Casualties Great Britain: 2014 Annual Report* (pp. 1–13).

Department for Transport. (2015). The pathway to driverless cars. In *The pathway to driverless cars: Summary report and action plan*. https://doi.org/10.1007/s13398-014-0173-7.2.

Designated Driver. (2019). World's first remotely-controlled 5g car to make history at goodwood festival of speed. Retrieved October 27, 2020, from https://designateddriver.ai/2019/07/worlds-first-remotely-controlled-5g-car-to-make-history-at-goodwood-festival-of-speed/.

Drury, J., Riek, L., & Nathan, R. (2006). A decomposition of UAV-related situation awareness. In *Proceeding of the 1st ACM SIGCHI/SIGART conference on human-robot interaction—HRI '06*, 88. https://doi.org/10.1145/1121241.1121258.

Drury, J. L., Keyes, B., & Yanco, H. A. (2007). *LASSOing HRI : Analyzing situation awareness in map-centric and video-centric interfaces* (pp. 279–286).

Endsley, M., Bolte, B., & Jones, D. (2003). *Designing for situation awareness* (2nd ed.). Boca Raton: CRC Press.

Endsley, M. R. (2020). Situation awareness in driving. In D. Fisher, W. J. Horrey, J. D. Lee, & M. Regan (Eds.), *Handbook of human factors for automated, connected and intelligent vehicles*. London: Taylor and Francis.

Endsley, M. R. (2012). *Designing for situation awareness: An approach to user-centered design*. Boca Raton: CRC Press. https://doi.org/10.1201/9780203485088.

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making, 9*(1), 4–32. https://doi.org/10.1177/1555343415572631.

Endsley, M. R., et al. (2019). Situation awareness in future autonomous vehicles: Beware of the unexpected. In S. Bagnara (Ed.), *Advances in intelligent systems and computing* (pp. 303–309). Cham: Springer. https://doi.org/10.1007/978-3-319-96071-5_32.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society Annual Meeting*. https://doi.org/10.1177/154193128803200221.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Aerospace and Electronics Conference*. https://doi.org/10.1109/NAECON.1988.195097.

Endsley, M. R. (2017). Autonomous driving systems: a preliminary naturalistic study of the tesla model S. *Journal of Cognitive Engineering and Decision Making*. https://doi.org/10.1177/1555343417695197.

Endsley, M. R. (2017). From here to autonomy : lessons learned from human—automation research. *Human Factors, 59*(1), 5–27. https://doi.org/10.1177/0018720816681350.

Ericsson. (2017). *Remote operation of vehicles with 5G (extract from the Ericsson mobility report)* (June). Retrieved from https://www.ericsson.com/assets/local/mobility-report/documents/2017/emr-november-2017-remote-operation-of-vehicles-with-5g.pdf.

Fisher, D. L., Pradhan, A. K., Pollatsek, A., & Knodler, M. A. (2007). Empirical evaluation of hazard anticipation behaviors in the field and on driving simulator using eye tracker. *Transportation Research Record, 2018,* 80–86. https://doi.org/10.3141/2018-11.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). Take over! How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931213571433.

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations. *Human Factors, 58*(4), 642–652. https://doi.org/10.1177/0018720816634226.

Gugerty, L. J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*(1), 42–66. https://doi.org/10.1037/1076-898X.3.1.42.

Gugerty, L. J. (2011). Situation awareness in driving. *Handbook for Driving Simulation in Engineering, Medicine and Psychology*. https://doi.org/10.1518/001872008X288394.

Hampshire, R. C., Bao, S., Lasecki, W. S., Daw, A., & Pender, J. (2020). Beyond safety drivers: Applying air traffic control principles to support the deployment of driverless vehicles. *PLoS ONE, 15*(5), 1–15. https://doi.org/10.1371/journal.pone.0232837.

Hancock, P. A., Nourbakhsh, I., & Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences of the United States of America, 116*(16), 7684–7691. https://doi.org/10.1073/pnas.1805770115.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 904–908).

Hawkins, A. J. (2020). Exclusive look at cruise's first driverless car without a steering wheel or pedals. Retrieved from https://www.theverge.com/2020/1/21/21075977/cruise-driverless-car-gm-no-steering-wheel-pedals-ev-exclusive-first-look.

Hillman, R., & Capaldi, R. (2020). Test methods for interrogating autonomous vehicle behaviour. Retrieved from HORIBA MIRA website: https://register.gotowebinar.com/recording/4054104289357790475.

Hosseini, A., & Lienkamp, M. (2016). Enhancing telepresence during the teleoperation of road vehicles using HMD-based mixed reality. In *IEEE intelligent vehicles symposium, proceedings*, *2016-Augus*(Iv) (pp. 1366–1373). https://doi.org/10.1109/IVS.2016.7535568.

Jones, D., & Endsley, M. R. (1996). Sources of situation awareness errors in aviaiton domain. *Aviation Space and Environmental Medicine, 67*(6), 507–512. https://doi.org/10.1039/c4qo00187g.

Kemp, I. (2018). *Autonomy & motor insurance what happens next ? An RSA report into autonomous vehicles & experiences from the GATEway Project* (July) (pp. 1–31).

Khattak, Z. H., Fontaine, M. D., & Smith, B. L. (2020). Exploratory investigation of disengagements and crashes in autonomous vehicles under mixed traffic: An endogenous switching regime framework. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/tits.2020.3003527.

Koopman, P., & Wagner, M. (2016). Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety, 4*(1), 15–24. https://doi.org/10.4271/2016-01-0128.

Koopman, P., & Wagner, M. (2017). Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine, 9*(1), 90–96. https://doi.org/10.1109/MITS.2016.2583491.

Lo, J. C., Sehic, E., Brookhuis, K. A., & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation Research Part F: Traffic Psychology and Behaviour, 43,* 325–338. https://doi.org/10.1016/j.trf.2016.09.006.

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931214581351.

Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics, 60,* 293–304. https://doi.org/10.1016/j.apergo.2016.12.003.

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics, 35*(10), 939–953. https://doi.org/10.1016/j.ergon.2005.04.002.

Manus. (2020). Škoda: The ultimate driving experience. Retrieved November 16, 2020, from https://manus-vr.com/use-cases/skoda.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. *Procedia Manufacturing, 3,* 2867–2873. https://doi.org/10.1016/j.promfg.2015.07.788.

Mills, J. (2018). Tesla owner filmed in passenger seat using autopilot says he was "unlucky one who got caught." *The Sunday Times*. Retrieved from https://www.driving.co.uk/video/tesla-owner-filmed-passenger-seat-using-autopilot-says-unlucky-one-got-caught/

Mok, B., Johns, M., Miller, D., & Ju, W. (2017). Tunneled in: Drivers with active secondary tasks need more time to transition from automation. In *Proceedings of the 2017 CHI conference on human factors in computing systems—CHI '17* (pp. 2840–2844). https://doi.org/10.1145/3025453.3025713.

Move_UK. (2018). *Data analysis report phase 2* (November).

NASA. (2017). NASA, Boeing Testing Synthetic Vision Technologies.

NHTSA. (2016). The road ahead. Retrieved January 10, 2019, from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/12532-nhtsa-strategicplan-2016-2020.pdf.

Niklasson, L., Riveiro, M., Johansson, F., Dahlbom, A., Falkman, G., Ziemke, T., et al. (2007). A Unified Situation Analysis Model for Human and Machine Situation Awareness. *Herzog, O., Rödiger, K.-H., Ronthaler, M. & Koschke, R. (Hrsg.), Informatik 2007 – Informatik Trifft Logistik – Band 2. Bonn: Gesellschaft Für Informatik e. V.. (S. 105–109).*, *2*, 105–109.

NTSB. (2018). *Preliminary Report HWY18MH010*. 4. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf.

OFCOM. (2019). *Connected Nations 2019: UK report*. (December), 1–49. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0023/186413/Connected-Nations-2019-UK-final.pdf.

Ottesen, A. (2014). *Situation awareness in remote operation of autonomous ships* (pp. 1–12).

Oxbotica. (2020). Oxbotica: About us. Retrieved October 23, 2020, from https://www.oxbotica.com/about-us/

Pamungkas, D. S., & Ward, K. (2014). Electro-tactile feedback system for achieving embodiment in a tele-operated robot. In *2014 13th international conference on control automation robotics & vision (ICARCV)*, *2014*(December) (pp. 1706–1711). https://doi.org/10.1109/ICARCV.2014.7064573.

Pollard, T. (2018). What are autonomous car levels ? Levels 1 to 5 of driverless vehicle tech explained. Retrieved January 10, 2019, from https://www.carmagazine.co.uk/car-news/tech/autonomous-car-levels-different-driverless-technology-levels-explained/.

Porathe, T., Prison, J., & Man, Y. (2014). Situation awareness in remote control centres for unmanned ships. *Human Factors in Ship Design & Operation*, (February), 1–9.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the Human Factors and Ergonomics Society*, *2014-Janua*(1988) (pp. 2063–2067). https://doi.org/10.1177/1541931214581434.

Ricaud, B., Lietar, R., Joly, C., Ricaud, B., Lietar, R., Joly, C., et al. (2017). Are virtual reality headsets efficient for remote driving? In *Simulation 2015 (RSS 2015), Oct 2015, Orlando, United States. Proceedings of the international conference on road safety simulation.*

Ruano, S., Cuevas, C., Gallego, G., & García, N. (2017). Augmented reality tool for the situational awareness improvement of UAV operators. *Sensors (Switzerland), 17*(2), 1–16. https://doi.org/10.3390/s17020297.

Russell, D. M., & Dieberger, A. (2002). *Synthesizing evocative imagery through design patterns* (pp. 3–6).

SAE International. (2016). Automated driving: Levels of driving automation are defined in new SAE International Standard J3016. In *SAE International*.

SAE International. (2018). Surface vehicle. In *SAE International*.

Salmon, P. M., Stanton, N. A., & Young, K. L. (2012). Situation awareness on the road: Review, theoretical and methodological issues, and future directions. *Theoretical Issues in Ergonomics Science, 13*(4), 472–492. https://doi.org/10.1080/1463922X.2010.539289.

Scholtz, J., Antonishek, B., & Young, J. (2004). Operator interventions in autonomous off-road driving: effects of terrain. In *2004 IEEE international conference on systems, man and cybernetics (IEEE Cat. No.04CH37583)* (Vol. 3, pp. 2797–2802). https://doi.org/10.1109/ICSMC.2004.1400756.

Störmer, V. S. (2019). Orienting spatial attention to sounds enhances visual processing. *Current Opinion in Psychology, 29,* 193–198. https://doi.org/10.1016/j.copsyc.2019.03.010.

T Systems. (2020). *Teleoperated driving:Remote vehicle control*.

Tang, T., Kurkowski, J., & Lienkamp, M. (2013). Teleoperated road vehicles: A novel study on the effect of blur on speed perception. *International Journal of Advanced Robotic Systems*. https://doi.org/10.5772/56735.

Thorn, E., Kimmel, S., & Chaka, M. (2018). A framework for automated driving system testable cases and scenarios. In *National highway traffic safety administration*. Retrieved from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automateddrivingsystems_092618_v1a_tag.pdf.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522. https://doi.org/10.1038/381520a0.

Tibken, S. (2018). Waymo CEO: Autonomous cars won't ever be able to drive in all conditions. Retrieved from cnet.com website: https://www.cnet.

com/news/alphabet-google-waymo-ceo-john-krafcik-autonomous-cars-wont-ever-be-able-to-drive-in-all-conditions/.

UNECE. (2020). Human factors challenges of remote support and control: A position paper from HF-IRADS. *Informal Document No. 8—September 2020 session of WP.1*, (8) (pp. 1–9). Retrieved from https://www.unece.org/fileadmin/DAM/trans/doc/2020/wp29grva/GRVA-07-65e.pdf.

United Lex. (2017). Virtual Reality Powers Remote Driving. Retrieved October 21, 2018, from https://www.unitedlex.com/news-and-insights/blog/2017/virtual-reality-powers-remote-driving.

Waymo. (2020). *Waymo safety report our mission* (September).

Wevolver. . (2020). 2020 Autonomous Vehicle Technology Report. *Muma Business Review, 3,* 193–205.

Wood, M, Robbel, Maass, M, Tebbens, R, Meijs, M, et al. (2019). Safety first for automated driving. In *Aptiv Services US, LLC; AUDI AG; Bayrische Motoren Werke AG; Beijing Baidu Netcom Science Technology Co., Ltd; Continental Teves AG & Co oHG; Daimler AG; FCA US LLC; HERE Global B.V.; Infineon Technologies AG; Intel; Volkswagen AG.*

Yanco, H., & Drury, J. (2004). "Where am I?" Acquiring situation awareness using a remote robot platform. *Conference Proceedings IEEE International Conference on Systems, Man and Cybernetics, 3,* 2835–2840. https://doi.org/10.1109/ICSMC.2004.1400762.

Publisher's Note

Paper 2

Mutzenich, C, Helman, S, Durant, S & Dalton, P. (2021). Situation Awareness in Remote Operators of Autonomous Vehicles: Developing a Taxonomy of Situation Awareness in Video-Relays of Driving Scenes. *Frontiers in Psychology*, *12* (November). https://doi.org/10.3389/fpsyg.2021.727500.

frontiers in Psychology

ORIGINAL RESEARCH
published: 10 November 2021
doi: 10.3389/fpsyg.2021.727500

Situation Awareness in Remote Operators of Autonomous Vehicles: Developing a Taxonomy of Situation Awareness in Video-Relays of Driving Scenes

Clare Mutzenich[1], Szonya Durant[1], Shaun Helman[2] and Polly Dalton[1]*

[1] Royal Holloway, University of London, London, United Kingdom, [2] Transport Research Laboratory, Crowthorn, United Kingdom

Edited by:
Leila Takayama,
University of California, Santa Cruz,
United States

Reviewed by:
Nikolas Martelaro,
Carnegie Mellon University,
United States
Federica Caffaro,
Roma Tre University, Italy
George Yannis,
National Technical University
of Athens, Greece

***Correspondence:**
Clare Mutzenich
clare.mutzenich.2018@live.rhul.ac.uk

Specialty section:
This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 18 June 2021
Accepted: 05 October 2021
Published: 10 November 2021

Citation:
Mutzenich C, Durant S, Helman S
and Dalton P (2021) Situation
Awareness in Remote Operators
of Autonomous Vehicles: Developing
a Taxonomy of Situation Awareness
in Video-Relays of Driving Scenes.
Front. Psychol. 12:727500.
doi: 10.3389/fpsyg.2021.727500

Even entirely driverless vehicles will sometimes require remote human intervention. Existing SA frameworks do not acknowledge the significant human factors challenges unique to a driver in charge of a vehicle that they are not physically occupying. Remote operators will have to build up a mental model of the remote environment facilitated by monitor view and video feed. We took a novel approach to "freeze and probe" techniques to measure SA, employing a qualitative verbal elicitation task to uncover what people "see" in a remote scene when they are not constrained by rigid questioning. Participants ($n = 10$) watched eight videos of driving scenes randomized and counterbalanced across four road types (motorway, rural, residential and A road). Participants recorded spoken descriptions when each video stopped, detailing what was happening (SA Comprehension) and what could happen next (SA Prediction). Participant transcripts provided a rich catalog of verbal data reflecting clear interactions between different SA levels. This suggests that acquiring SA in remote scenes is a flexible and fluctuating process of combining comprehension and prediction globally rather than serially, in contrast to what has sometimes been implied by previous SA methodologies (Jones and Endsley, 1996; Endsley, 2000, 2017b). Inductive thematic analysis was used to categorize participants' responses into a taxonomy aimed at capturing the key elements of people's reported SA for videos of driving situations. We suggest that existing theories of SA need to be more sensitively applied to remote driving contexts such as remote operators of autonomous vehicles.

Keywords: situation awareness (SA), SA comprehension, SA Prediction, driving, video, taxonomy

INTRODUCTION

Recent years have seen a transition toward ever-increasing levels of vehicle autonomy. With the aim of providing a universal taxonomy for defining levels of automation, the organizational body SAE International (2016) highlights six levels of automation for on-road vehicles. At the final stage of this taxonomy the car is fully self-driving, and the occupant is never required to take over. Yet even entirely driverless vehicles of this type will sometimes require human intervention. For example, an autonomous vehicle (AV) is unlikely to be able to interpret a contractor directing traffic using hand signals at a construction site, whereas a human can do so easily. In "edge" cases such as this, it is likely that human operators will step in to interpret the unexpected situation, and their input will need to be provided remotely in vehicles with no "backup" driver present at the scene.

The established understanding of automated driving is being continuously updated and now recognizes the need for occasional remote operation even at the higher levels of automation (Mutzenich et al., 2021). The industry standard taxonomy, SAE J3016, which outlines and defines the modes of automated driving was further updated in April 2021 to include remote support functions (SAE International, 2021). It is likely that, in some circumstances, a remote operator (RO) may even need to take over the real-time driving of the car (for example, if the automation has completely failed). Our research sets out to address the significant human factors challenges unique to a remote operator temporarily in charge of an automated vehicle.

For ROs to build up a mental model of the environment that they do not physically occupy, they will need a rich array of information from a variety of different perspectives to help them. Most remote driving is facilitated by monitor view and video feed, which is limited in terms of the size of the view presented and the reliance on 2D depth cues, making it inevitably second-rate to being physically present inside the vehicle. Yet the differences in building and maintaining remote SA have historically received relatively little attention in a road transport context and certainly not in remote operators of automated vehicles. How SA is attained from video relay is an important consideration when designing remote operator interfaces and training programs.

An RO who has been alerted by an AV to "drop in" and assess the problem or assume direct driving control will first need to acquire situation awareness (SA) of the remote scene. SA encompasses what is known about the environment, what is happening in it and what might change. There are many definitions of SA (see Gugerty, 1997, 2011; Endsley et al., 2003; Niklasson et al., 2007; Endsley, 2015; Lo et al., 2016) but the most cited is from Endsley's model which divides knowledge of the environment into three levels of awareness: perception, comprehension and projection. However, the concept of projection (Level 3 SA) typically includes factual calculations not within the domain of the average driver (note that Endsley's original research used military pilots adept at interpreting data from multiple instruments). We argue that "prediction" is a better descriptor for this level of driver SA, as predictions are based on subjective analysis of known external factors, which may change and are uncontrollable, but nevertheless appear likely to happen on the basis of the driver's general experience. Therefore, we refer to the levels of SA simply as perception, comprehension and prediction (rather than projection).

Endsley's original SA model was developed using military pilots but attempts have been made to extend it to *in situ* driving contexts (Ma and Kaber, 2005; Bolstad et al., 2010; Endsley, 2019). Further models of SA have been developed which align more closely with the driving domain than Endsley's (Matthews et al., 2001). Matthews' SA model (2001) divided driving awareness into strategic, tactical, and operational goals, putting the emphasis on the driver's objectives. Strategic driving is ultimately goal orientated, such as destination and route, which may be part of existing mental models. Tactical driving is conscious and intentional, involving decisions such as when to overtake or change speed which require feedback from the environment and system. Operational driving implements tactical decisions into dynamic driving actions and maneuvers such as steering wheel control or braking. These driving goals interact to some extent and can also map onto Endsley's three levels of SA in varying degrees, for example strategic driving is mostly Level 3 Projection, but Level 1 Perception and Level 2 Comprehension are also integrated.

Research that has considered the functional role of SA in automation and driving has so far been limited to the SA level of an *in situ* driver who has to take over the driving of an AV that can drive alone for short periods of time, for example measuring the response times of drivers assuming command after adaptive cruise control has been employed (Banks et al., 2018); investigating the issues of complacency and overtrust when humans are required to be in an active monitoring state for prolonged periods of time in automated cars (Larsson et al., 2014) and the current trends in designing automated vehicles which influence *in situ* driver's SA (Walker et al., 2008). However, the prospect of ROs occasionally having to unexpectedly take control of AVs from a separate location makes it essential to identify how their SA needs will be different from those of an on-site driver.

Neither Endsley's nor Matthews' SA models tell us much about how people use the information in their environment to build SA. Some researchers have considered whether it is appropriate to apply Endsley (1995) model of SA to driving at all, querying whether it is more appropriate to refer to SA as cyclical, demonstrating top down and bottom processing rather than constructed *via* hierarchical levels (Neisser, 1976 as cited in Revell et al., 2020). Salmon et al. (2012) conducted a comprehensive review of SA and techniques and methodologies for assessing SA, arguing in favor of a systems-based explanation that considers the cycle of activity encompassing all road users together with the road environment and infrastructure, other vehicles and the impact of developing technologies combined. In which case, the question of how to effectively measure remote operator SA represents an operational challenge in the field.

Across the literature, the Situation Awareness Global Assessment Technique (SAGAT), a quantitative measurement of SA, is widely accepted to objectively measure SA (Endsley, 1988). SAGAT is a freeze and probe technique, where simulated trials are halted, and participants asked questions designed to measure SA levels of perception, comprehension and projection. Their responses are then compared with the reality of the simulation in order to derive a performance score. Previous attempts to measure SA in driving may lack construct validity as they are based on *a priori* notions of what driver SA "is" in the mind of the researcher. Studies that have adopted SAGAT to measure driving SA have typically measured people's awareness in a range of different ways e.g., using a different number of probes or trials (Scholtz et al., 2005; Ma and Kaber, 2007); using a reconstruction task or a recognition task (Gugerty, 1997; Franz et al., 2015); or using a verbal recording in real time (Endsley, 2017a).

Pre-determined questions are likely to artificially constrain the level of SA that participants are able to demonstrate (e.g., a person may be continually aware of the weather, but if they are not probed about this awareness, they will not be able to demonstrate this awareness) and, combined with the simulated nature and variability of the task set, do not get at the heart of what people "see" in their environment when viewing a remote driving scene. Nevertheless, we acknowledge Endsley, (2015, p. 9) argument that SA is part of "*deeply embedded mental models and schemas,*" and that the quantitative methodology that SAGAT adopts is necessary to extract information that people would otherwise be unable to communicate because they cannot be relied upon to have insight into their own SA.

Outline of the Current Study

We contend that it seems likely that some combination of qualitative and quantitative approaches will eventually provide the fullest understanding of remote driving SA. However, as a starting point perhaps the most direct technique to get at this abstract and conceptual data is simply to ask **what** people see in a remote scene and capture and interpret every detail of what they report. Given that remote driving SA is such a new field, it makes sense to start from an unconstrained tool to inspect the implicit and explicit mental processes that are underway while ROs build up a remote model of the environment. With this aim, the current study made use of verbal elicitation techniques to uncover, in its most basic form, what people "see" in a remote scene when they are not constrained by rigid questioning. Video elicitation methods which require videos to be narrated by an expert as to their thinking, decision making and interpretation of the stimuli, have been found to be helpful in identifying "invisible phenomena" that are difficult to abstract using quantitative methods (Jewitt, 2012, p. 4). Likewise, retrospective verbal protocol techniques which encourage participants to provide a commentary of information they drew from the environment and what they were thinking about it have been shown to measure the cognitive processes that support SA Perception and SA Comprehension and the feedback process that shapes SA Prediction (Walker et al., 2008). ROs will need to take decisions based on the SA that they have conscious access to, so freely elicited verbal description seems just as likely to get at deeply embedded mental models and schemas as a SAGAT probe.

Operators will have to build up a mental model of the remote environment facilitated by monitor view and video feed, meaning the task of developing SA from a remote location is likely to be made more difficult by the operator's physical absence, as well as any signal degradation. A comprehensive inventory of the mental models that underpin the construction of driving SA from video feeds is thus a clear research priority as video will surely play a role in remote operation. In this study, we used real footage filmed from a driving perspective, giving a naturalistic experience of a remote driving situation. We investigated how participants build up mental representations of these naturalistic remote driving scenes, using a verbal elicitation protocol at the end of the video. We also examined whether providing extra information from the rear-view camera footage influenced this process. Inductive thematic analysis was used to encode participants' responses into a taxonomy aimed at capturing the key elements of people's reported SA for videos of driving situations.

Justification of Method

Inductive thematic analysis (TA) is the practice of encoding qualitative data by searching across the whole raw data set to find recurring patterns, regarded as "themes," which are then used to generate a theory (Braun and Clarke, 2013). This type of analysis is likely to be less affected by researchers' *a priori* preconceptions of SA than deductive TA which searches for evidence in datasets of pre-existing theories (e.g., in the context of our research, Endsley's three "levels" of SA would have been the most obvious candidate for a pre-existing theoretical framework; (Boyatzis, 1998)). This paper adopts an inductive TA methodology to investigate more freely the complex processes that generate SA construction in remote viewing, which are not presently identified in quantitative metrics, generating deeper understanding of the semantic information that people can access in naturalistic driving scenes.

The inductive TA procedure in this paper followed the phases of thematic analysis outlined in Braun and Clarke (2013). Firstly, the lead researcher transcribed all participant verbal responses to familiarize herself with the data, then initial codes were generated using a data driven approach by listing all words used by each participant. Themes were developed using a semantic approach which reports the explicit meaning of codes, rather than a latent analysis which seeks to identify underlying meanings or assumptions behind the codes (Boyatzis, 1998). Themes were then grouped into patterns and reviewed by checking that they were mutually exclusive, and that the entire data set could be classified into the suggested themes. Finally, themes were named and defined according to how they best described the features of that theme. During the process, we assumed an active role in determining themes, reflected in our analysis where, rather than reporting themes as passively "emerging" from the data, which has been the subject of much criticism in this methodology, we are careful to provide sufficient detail in the process of determining the coding decisions that were made in respect to item definition and inclusion in the construction of themes (Braun and Clarke, 2013, p. 80).

The analysis of participant transcripts using this approach enabled the construction of a taxonomy of SA in video-relays of driving scenes. Taxonomies serve to classify concepts and give insight into the principles which underlie these classifications. Previous taxonomies relating specifically to driving have mainly focused on driving errors and violations to identify causal factors and risks involved on different highway and environmental contexts (Stanton and Salmon, 2009; Khattak et al., 2021) or in highly automated cars, the factors that lead to handovers in critical situations (Capallera et al., 2019) whereas our taxonomy delineates the mental models that underpin the construction of driving SA *via* video, facilitating an understanding of the themes that make up someone's remote SA. Once validated, it is anticipated that this taxonomy can be used to develop regulatory frameworks for training remote operators of AVs and will be used in future empirical work in our laboratory to design quantitative queries that can effectively measure the SA of ROs.

MATERIALS AND METHODS

Developing the Driving Video Stimuli

To create the driving videos, researchers drove a 2016 petrol Dacai Duster 4 × 4 (manual right-hand drive) for 32 min filming continuously. The 21.8 miles route was from Royal Holloway, University of London, Egham, using the following roads, A328/B329/A30/M3/A331/A30. The route was designed to encounter a range of roads covering all speed limits and including motorways, A roads, minor country roads and residential areas (see **Figure 1** for examples of different road types). This was in line with previous research that had created videos of on-road driving scenes (Walker et al., 2008; Salmon et al., 2013). This paper aims to address the lack of standardization in measuring driving SA by providing a complete, open-source set of remote videos of driving scenes[1].

Video and audio recordings of the forward and rear views were filmed using two GoPro6 cameras mounted using the GoPro Suction Cup Mount. The front camera was mounted 60 cm from the end of the bonnet and 50 cm lateral from the widest point to ensure that it was not obstructing the driver's view and the end of the bonnet was not in shot. The rear camera was mounted 20 cm up from the bottom of the rear-view window and 50 cm lateral position pointing at the road behind the vehicle. This distance was selected because it did not obstruct the driver's view and the body of the car was not visible. Both camera angles were verified from the driver's position inside the car to confirm they accurately reflected the available forward facing and rear-view perspective. This gave a first-person perspective of the road ahead and pilot tests verified that it accurately reflected what the driver was seeing using in-car footage relayed to a mobile phone. Both cameras were controlled inside the car and recording started as soon as the driver released the handbrake.

The 32-min recording was edited using Camtasia©[2] to create eight separate video segments divided into "A roads" (A30/331), "Residential roads" (A328), "Rural roads" (B329) and "Motorway roads" (M3). Two versions of each video were created; a forward-facing video recording of the total drive ("rear-view absent" condition) and a second version of the video with the addition of the rear facing footage positioned in the top left corner of the video ("rear-view present" condition) (see **Figure 2** for images showing the two versions of the video). In this study, rear-view footage of the car was presented to the left of the center of the driving video to increase mundane realism, as it is typically the location of the rear-view mirror in a car in the United Kingdom (Buscher et al., 2009). The image was set at a ratio of 34.8% of the total image in the top left corner of the screen. This size was based on the approximate proportional scale of a rear-view mirror to the windscreen in a car, as there are no standard sizes for either mirror or windshield in United Kingdom regulations.

We used the Gorilla Experiment Builder[3] to create and host our experiment (Anwyl-Irvine et al., 2018). Data was collected between 9th September 2019 and 7th October 2019. Participants were recruited through Prolific. Participants took part *via* a laptop or desktop computer, and we excluded the use of smartphones and tablets. Participants were advised that they must be in a quiet environment with no background noise or distractions. They were required to have a working microphone on their PC and completed a microphone audio test to proceed in the experiment. Videos were presented at 60 frames per second on an 811 × 456 pixel screen placement holder.

A preliminary study was conducted ($n = 10$) to test the audio recording task in Gorilla and check the quality of the driving videos during online testing. During this testing procedure, it was evident that participants were unsure of what to say and there were large differences in the amount and quality of content that they gave in response to the questions with some comments not directly relevant to driving. Other research that has used qualitative "think aloud" methodologies or verbal protocols has provided participants with practice trials and/or suggested the kinds of items or events that could be appropriate to mention (Walker et al., 2008; Salmon et al., 2013). This could be viewed as pushing participants to say certain kinds of things and could therefore introduce bias. However, the limited range and detail of observations in the pilot transcripts suggested that an example trial was necessary in the current experiment, to ensure that participants gave sufficiently full accounts. For this reason, the final version of the experiment included an example trial, described in more detail in the "Procedure" section below.

During the preliminary study, we discovered that the road sounds that accompanied the videos were likely to provide extra situation awareness cues to aspects such as speed, possibly creating a more immersive experience. Some remote operation training teaches operators how to pick up cues from other feedback, for example spatial audio above, around and behind the car (Tang et al., 2013; Störmer, 2019; UNECE, 2020). Although participants in this study were unlikely to possess this type of skill, they may not have all had audio enabled on their computers, so we removed the audio track from the videos to reduce this source of variability between the experimental stimuli. The effects of the presence (vs. absence) of this audio is an interesting issue for future empirical studies, and some of this research is currently underway in our lab.

Design

We adapted the SAGAT methodology to create a qualitative task. At the end of each video, participants were asked to respond verbally to two open questions: "what is happening?" (to determine their comprehension of the scene); and "what will happen next?" (assessing the ability to make predictions from their understanding of the scene).

We also examined whether providing extra information from the rear-view camera footage influenced this process. In a within groups experimental design, all participants watched four videos with a rear-view mirror present ("rear-view present") and four videos without a rear-view mirror ("rear-view absent") in a random order. Presentation was counterbalanced with respect to road type. For example, if Rural #1 was viewed "rear present" the rear-view footage, the other version of that road type (Rural #2)

---

[1] https://github.com/clareyclare/clareyclare/blob/main/video-relays%20of%20driving%20scenes

[2] www.techsmith.com, Version 2018. 0. 6.

[3] www.gorilla.sc

FIGURE 1 | Examples of different road types: **(A)** motorways, **(B)** a roads, **(C)** minor country roads, and **(D)** residential areas.

FIGURE 2 | Images showing the two versions of the video, without (**left** image) and with a rear-view insert (**right** hand image).

was presented "rear absent." This was to eliminate any influence of the presence of the rear-view footage potentially affecting performance differentially across road types. We expected that the combination of both fields of view ("rear-view present" condition) would enhance SA by giving a more immersive experience of the remote scene.

Participants

Ten participants were recruited randomly online using Prolific (an online participant recruitment tool[4]), meaning the sample was split unevenly by gender, with three quarters (70%) of the participants being female. All participants were United Kingdom residents with English as their first language and possessed a full United Kingdom driving license for a minimum of 3 years. We used bracketed ranges in increments of five to ask participants their age, the number of years they had held a driver's license and we asked them how frequently they drove (daily, weekly, monthly). We also asked them to record their approximate annual mileage (in 1000 km). **Table 1** shows a summary of participant demographic details. Half of the participants drove on a daily basis and only one participant rarely drove. An upper age limit of 75 was set to coincide with United Kingdom driving laws but, during testing, the maximum age bracket selected was 61–65 years old. There were two modal age ranges 31–35 and 41–45. Seven participants drove 5, 000 miles or more in the last year and had 10 years or more driving experience.

Procedure

Participants were informed that they would see videos of driving scenes and would be asked two questions about the last few seconds of the driving scene. They were instructed to consider themselves the driver and to describe the road that "you" were on, driving maneuvers "you" were carrying out or the behavior of other road users and pedestrians. For the SA Prediction

---

[4] www.prolific.co

TABLE 1 | Participant demographic Information (*n* = 10).

Participant number	Gender	Age	Annual mileage (in 1000 km)	Years held driver's license	Frequency driving
1	Male	41–45	5–10	21–25	Weekly
2	Female	21–25	<1	<3	Rarely
3	Male	31–35	10–15	11–15	Weekly
4	Female	41–45	1–5	21–25	Daily
5	Female	21–25	5–10	3–5	Weekly
6	Female	61–65	5–10	46–50	Daily
7	Female	26–30	5–10	6–10	Daily
8	Female	31–35	>20	11–15	Daily
9	Female	41–45	<1	21–25	Daily
10	Male	31–35	5–10	11–15	Weekly

questions, they were advised that they could concentrate on possible future directions "you" may take, the actions of other drivers or road users or the changing physical environment around "you." We reminded participants to press the "Start recording" button and "Stop recording" button to record their answers after each video. We showed participants an example video (30 s) and played them example spoken answers, recorded by a confederate, in response to the two experimental questions. The example video was from the "rear absent" condition so that they were not primed to the nature of the independent variable before the experimental trials started. **Table 2** shows the scripts of the recorded example audio for SA Comprehension and SA Prediction.

Participants then watched 8 stimulus videos one after another, responding to the two SA questions and the video quality measure after each video (see **Figure 3** for a schematic summarizing the procedure). In the debrief, participants were asked "*Do you have any comments about your experience of being a participant in this experiment? For example, was anything unusual or gave you difficultly while you were carrying out the study?*" which was to collect data regarding their viewing experience of the videos.

Inductive Thematic Analysis of Participant Transcripts

Participants' verbal responses were recorded as mp3 audio files in Gorilla. The audio files were downloaded into NVivo qualitative data analysis software (QSR International Pty Ltd., Version 11, 2015) and were transcribed by the lead researcher. No participants indicated quality problems that required their results to be excluded.

SA Comprehension and SA Prediction question responses were analyzed separately using the same analysis procedure. All ten participant recordings for each video were analyzed together as one set. An inductive thematic analysis procedure was used to evaluate participants' naturalistic situation awareness to produce a taxonomy of situation awareness in driving.

Creating classifications to represent driving SA involved extracting information elements from participant transcripts and establishing shared temporal, spatial and semantic concepts between them using inductive thematic analysis. Researchers divided each transcript into singular words and coded each word as an individual item. Similar words or concepts verbalized by participants were grouped together and recorded as sub-themes; for example, the sub-theme "Type of highway" contained the items, "country lane/a country road/a country road/country lane/a rural road/down a country road/road."

A fundamental objective in coding qualitative responses to stimuli is to make the decisions involved in developing the themes transparent and clearly linked to the aim (Graneheim et al., 2017). The research team, consisting of four people, conducted blind tests to validate the decisions made as to what constituted an item. Each person judged the same five, randomly selected, participant transcripts from different roads and identified the items that were reported in each road transcript. Refinements to the coding scheme were discussed and agreed for each iteration until all researchers were within one item agreement on a final blind test.

Reliability Assessment

To judge whether the taxonomy was a robust measure of driving SA that could be used by other researchers, inter rater reliability was assessed. Inter rater reliability establishes the degree to which observers agree on the occurrence of each coded item, in this case, into the separate themes of the taxonomy (Jansen et al., 2003).

To collect a representative sample of the data set, 32 transcripts (16 for each of SA Comprehension and SA Prediction responses to the stimulus questions) across all eight video/road types, signifying 20% of the total dataset, were selected. All road types (8) were represented twice in random order and participant transcripts were randomly drawn with no replacement until all ten participants had been sampled equally, repeating the procedure until 16 transcripts had been selected.

An independent coder was trained in the coding of the taxonomy and their tallied totals for each theme were compared with the lead researcher's to calculate inter-rater reliability. We adopted a consensus approach to coding, meaning that if raters disagreed they would discuss how to apply the rating scale in that instance, allowing them to come to a decision on how to deal with the conflicting scores (Stemler, 2004).

Cohen's kappa was calculated for each SA Comprehension theme (6) and SA Prediction theme (4). Cohen's kappa was appropriate as it controls for chance agreement where raters

TABLE 2 | Example audio played to participants during the practice trial.

SA Question	Confederate transcript
Comprehension 'What was happening when the video stopped?'	*I'm on a residential road with houses... on both sides and trees..., on both sides... I just went past a big driveway on the left hand side with gates and two ...red... cars went past me on the right hand side. I'm just coming up to a signpost ...and there is a cyclist on the opposite side of the road coming towards me.*
Prediction 'What will happen next?'	*"The road in front of me was... straight and didn't have any turns... so I will carry on driving... straight ahead at the same speed... which was 30 mph ...and the cyclist will cycle past me on the other side... right hand side of the road."*

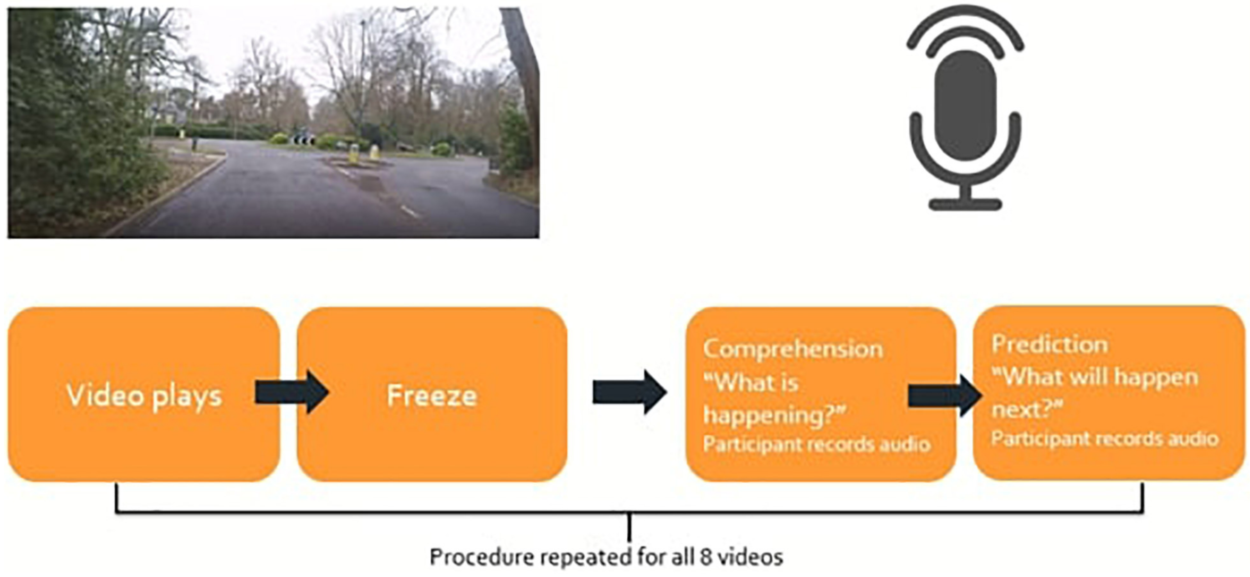


FIGURE 3 | Experimental procedure.

may be making random guesses if they are not sure which theme an item may fall into. There is general unanimity in the literature that any kappa below 0.60 indicates inadequate agreement (McHugh, 2012). At time 1 analysis, inter-rater reliability between the two raters for each theme was between fair (20–40%) and moderate (40–60%) which represents low agreement and could invalidate the usability of the taxonomy.

On review of the coding tables for each rater, it was found that one tally placed in the "wrong" theme could push two or more themes out of agreement so careful analysis was carried out to discover where the discordance occurred. There were several decisions taken when coding the transcripts as to what constituted an "item"; some verbalizations were ignored, only counted once or moved to a different SA response. The following subsections outline the coding decisions that were made in respect to item definition and inclusion.

Treatment of Errors

Two participants (P2 and P6) each had one inaccuracy in their reporting. P2 cited a second roundabout that was not present in the video and P6 claimed that a car had "*flashed lights at me*" when it had not. These errors were not included in the total item list. Similarly, if participants said something that they couldn't know, as it was not included in the video, we did not include it in the item list, for example,

> "*...I waited my turn and then proceeded to indicate and go round the vehicle.*" Participant 3

P3 cannot know whether the "driver" indicated or not as this was not shown in the video, so this is not a true SA observation but instead an artifact of schema driven expectations (we "should" indicate when we go round a vehicle). Schemas are discussed in detail in section "Role of Context in Comprehension Illustrated by Use of Schemas."

Coding Repetitions in Participant Transcripts

If participants referred to the same "item" more than once for example a pedestrian but in different contexts, we decided that it should be coded semantically, for example:

> "*... so I've pulled in to let other cars pass me... meanwhile the **young lady** is **swinging a bag** rather haphazardly **has walked past me** on the path and walked **further up the road** and then... as I finally manage to drive past the parked cars... **she's there** in front of me **swinging her bag.***" Participant 8

The reference to firstly, the *presence* of the pedestrian and their *location* ("walked past me") is recorded but also their *actions* ("swinging a bag"), *age* ("young") and *gender* ("lady") of that pedestrian were each counted as distinct items rather than subsuming these details into a generic "pedestrian." To gain a full sense of how participants view a driving scene from a remote perspective, every part of their description indicated how they were building up a mental model of the scene and so was included in our analysis. However, repeated mentions of identical information (e.g., the second mention in this example that the person was swinging her bag) were not coded separately.

In the SA Prediction question, some participants suggested more than one possible outcome based on what they had just seen, for example,

> "*…there will be either a roundabout or…a turn in the road.*" Participant 2.

It was decided to count each of these possible outcomes as valid, because evaluation of the future may produce multiple hypotheses.

Irrelevant Commentary in Participant Transcripts

On occasion, participants did not appear to fully put themselves in the ego perspective of the "driver" in the video, commenting that the driving did not reflect what they would have done in a particular situation, for example:

> "*I probably waited longer than normal, I would have overtaken and not waited as long normally*" Participant 4

> "*… I went from the slow lane to the middle lane… for no apparent reason because there is nothing in the slow lane for me to overtake and you should only be in the middle lane when you are overtaking…*" Participant 8

As these were not SA observations, these opinions were not coded as items, although the roadway descriptions within them ("slow lane," "middle lane") were included in the sub-themes. However, they did illustrate important distinctions between participants and how immersive they found the experiment, which we discuss in further detail in the results section "Understanding of Remote Vehicle State."

Relationships Between SA Comprehension and SA Prediction Transcripts

In some responses, participants did not adhere to the stimulus question and included information that was out of place. Some participants merged SA Comprehension with SA Prediction by referring to future actions (prediction – shown here in bold) in the first question (comprehension – shown here unbolded).

> *"I am on a bend turning right and there are cars passing me on the right. To go on the bend, I* ***will be slowing down my speed and being careful to make sure that. be sure because I don't have much visibility going around the bend.*** *"* Participant 2

To develop distinct themes in the taxonomy between SA Comprehension and SA Prediction, we removed any references to future actions from the item list for the comprehension category but ensured that they were still included in the SA Prediction category. In the responses to the SA Prediction question, it was agreed that only information in the future tense (as a prediction) would be counted.

Item Categorization in SA Prediction Question

It is difficult to make definitive judgments concerning what constitute "reasonable" predictions in driving situations. There are an infinite number of "items" that could be mentioned which make any classification system non-exhaustive. To recognize this, we differentiated between "abstract risks," which could be any conceivable event and "specific risks" which were named risks that directly related to the events unfolding in the video, as sub-themes within the over-arching theme of "Impending Hazards."

Also, it is not clear how to account for predictions concerning the absence of items or events, for example:

> "*I actually don't think anything will happen.*" Participant 6

This comment indicates a significant analysis of the absence of risk and would presumably be acted upon accordingly in a real driving situation, however, there is no noun which can be coded as an item. We decided to record this type of response under "Absence of hazards" e.g., no bends ahead/no signs/no pedestrians/no cars present/road clear/nothing will happen.

To re-assess inter-rater reliability of the taxonomy in light of all of these decisions, both the lead researcher and the independent observer were re-trained in the taxonomy. All transcripts were re-coded by both raters, with time 2 analysis exceeding the 80% threshold for almost perfect agreement, producing greater than 85% agreement in all SA Comprehension themes and between 80% and 91% for all SA Prediction themes (McHugh, 2012).

RESULTS

Development of a Taxonomy of Situation Awareness in Driving

The sub-themes produced by coding at item level across all road types were pooled to produce a complete list of the items that were mentioned in response to the SA Comprehension and SA Prediction questions. These coded items were grouped into common sub-themes, for example relating to the road/highway or relevant to the driver's perspective. These sub-themes were collated into over-arching themes which described systematic SA concepts that underpin naturalistic viewing of driving videos. There were 6 over-arching themes in SA Comprehension and 4 over-arching themes in SA Prediction (see **Table 3**).

The themes, together with the sub-themes within each, were recorded in a taxonomy of situation awareness in driving which encompasses the full range of what participants told us they saw in the videos of the driving scenes. The total proportion of references to each theme in the whole dataset were calculated and can be seen in **Table 4**, together with the frequency (*n*i) at which each of the sub-themes were referenced within each theme. It is important to note that these frequencies will have been influenced by the specifics of the scenes presented in the videos. Nevertheless, because a wide range of road types and scenarios were represented in the videos, and because many aspects were present in all videos (e.g., the road/highway and the driver's perspective) we believe that the frequencies of mention for each theme and sub-theme can be informative as long as they are interpreted with caution. The next two sections discuss each of the themes and some examples of sub-themes of the taxonomy. Further examples of each theme with quotes from the transcripts are illustrated in **Table 5**.

TABLE 3 | Taxonomy of situation awareness in video-relays of driving scenes for SA Comprehension and SA Prediction.

SA Comprehension

1.1 Road/highway

- ***Type of highway***
  - E.g., country road/lane/main road/city road/dual carriageway/motorway/A road/B road
  - Named road
  - E.g., M3, A30
- ***Road description***
  - E.g., two lanes/single lanes
  - E.g., busy/quiet/long/clear
  - E.g., speed limit
- ***Type of area***
  - E.g., village/residential
- ***Road layout information***
  - Turn
  - Straight road
  - Junction
  - Bends
  - Uphill/downhill
  - Roundabout (exits 1/2/3)
  - Roadworks/diversions/road blocked
  - Cycle lane
  - Traffic lights
  - Filter lanes
  - Slip road
  - Location of road layout information
  - To the left/right/middle of the road
- ***Road layout (adjective)***
  - "sharp" bend, "slight" bend, "very big roundabout" etc.,
- ***Time of day/season***
  - E.g., dusk/dark winter's day
- ***Traffic signs and signals***
  - On road signs and markings
  - Presence of sign/markings
  - Meaning of sign/markings
  - Location of sign/markings
  - Signpost
  - Presence of sign
  - Meaning of sign
  - Location of sign
  - Light signals
  - Traffic lights (red/amber/green)
  - Level crossings
  - Motorway signals e.g., lane control, fog, information message, camera
- ***Road "decoration"***
  - Trees
  - Houses
  - Location of "decoration"
  - Left hand side of the road (driver's side)
  - Right hand side of the road (offside)
  - Above (hanging)

1.2 Driver's perspective

- ***Relative position to other cars***
  - Behind
  - To the right
  - To the left
  - In front/driver following
  - Oncoming traffic on opposite side of the road
  - Passing parked cars
  - Distance e.g., close/far away
  - No cars present
  - No cars behind
- ***Direction/trajectory of driver***
  - Going right/left/straight on
  - Position on the road e.g., middle lane, second lane on motorway
- ***Speed traveling at***
  - Actual (mph)
- ***Subjective judgment of speed***
  - Fast/too fast/slowly
- ***Driver's decision/judgment of own action/s***
  - Temporal judgment (e.g., whether time to pass parked cars)
  - Decision to wait
  - Decision to maneuver out/overtake/nudge' into oncoming traffic
  - Should slow down
- ***Rationale of judgment of***
  - own action/s
  - Significance of obstacles (not enough space for two cars to fit)
  - Why driver is waiting/stationary (waiting for a gap in traffic)
- ***Driver perception issues***
  - Lack of visibility
  - Poor light

1.3 Other drivers/vehicles

- ***Description***
  - "Vehicle" or type of vehicle (e.g., van/lorry/car etc.,)
  - Many vehicles e.g., "traffic"
  - Make
  - Model
  - Color
  - Registration number
  - Body details (e.g., cages on the side)
  - Identification of company logo (e.g., DPD)
- ***Direction/trajectory of moving vehicle***
  - Going right/left/straight on/in left lane
- ***Speed traveling at***
  - Actual (mph)
  - Perceived/subjective e.g., "too fast"
  - At the speed limit
  - Queuing traffic/congested
  - Traffic flowing freely
- ***Purpose of vehicle***
  - E.g., Delivery/courier, waste/rubbish truck, local authority vehicle, ambulance
- ***Location of static vehicle/s***
  - On the verge/On the pavement/Left side/Right side
- ***State of vehicle***
  - Parked
  - Abandoned
- ***Rationale/analysis of other driver action (theory of mind)***
  - Why driver is slowing down (e.g., driver is looking for an address)
  - Why driver is parked (e.g., driver is going to deliver a package, doing ground works, clearing debris)

(Continued)

TABLE 3 | (Continued)

SA Comprehension		
1.4 Dynamic actions of driver and other drivers/vehicles	**1.5 Pedestrians/other road users**	**1.6 Anticipatory Hazards**
Maneuver planning/positioning Approaching bend Approaching roundabout Approaching traffic lights Waiting Approaching exits Signaling ***Active maneuvers*** Driving straight/Continuing driving Slowing down/speeding up/braking/stopping Going round bend Passing junction/turning Entering/Exiting roundabout (exits 1/2/3) Pulling out/Changing lanes/Pull back into lane Entering roads/motorway Taking exits/coming off Overtaking cars Give way ***Observation*** Checking mirrors Looking for other vehicles Observing other vehicle's signals (flashing lights, hazards etc.,)	***Presence of pedestrian/other road user*** E.g., "pedestrian," "cyclist," "rider," "dog" ***Location of pedestrian/other road user*** On pavement On side of road ***Location of pedestrian/other road user relative to driver*** On my left On my right Right hand side Left hand side On the other side of the road ***Description of pedestrian/other road user*** Outfit/clothes wearing (e.g., jacket) Color of clothes Holding or carrying specific items (e.g., bag) ***Gender of pedestrian/other road user*** Male/man/boy Female/woman/lady/girl ***Age of pedestrian/other road user*** Young Old ***Action of pedestrian/other road user (what pedestrian is doing)*** E.g., walking. swinging bag ***Purpose of action of pedestrian/other road user (why they are doing it)*** Theory of mind projection E.g., pedestrian stepping into road without looking/driver getting out of car into road without checking for oncoming traffic ***Occupation of pedestrian/other road user*** E.g., local authority worker	***Keeping distance from other vehicles*** E.g., not getting too close, keeping a car length between own car and car in front, being "careful" ***Watching out for other vehicles' behavior*** E.g., caution before pulling out on a roundabout
2. SA Prediction		
2.1 Changing road environment	**2.2 Driver future dynamic action/s**	
Road layout perception E.g., bend to the right/road to the left/4 lane motorway ***Environmental markers*** E.g., /railway/hump back bridge/houses/trees ***Traffic signs ahead*** On road e.g., road marking/mph/slow) Sign post e.g., roundabout/motorway/brown sign) Traffic lights (on red/amber/green) ***Changing road details*** E.g., Approaching roundabout/bend approaching/road turning to the right/going uphill/changing traffic lights	***Driver future observation*** Checking mirrors Looking for other vehicles Observing other vehicle's signals (flashing lights, hazards etc.,) ***Driver future maneuvers*** E.g., Overtake/turn/accelerate/continue to drive/stop/pull out/wait/drive on other side of the road/signal/give way to the right or left/finish maneuvers already started ***Driver future orientation/trajectory*** E.g., Straight on/round the bend/carry on/turn left or right ***Driver future speed*** E.g., mph/faster/slow down/slow down/maintain speed ***Driver position on the road*** E.g., in my lane/on the left hand side of the road/on a bend ***Reaching destination*** E.g., arriving at intended location	

(Continued)

TABLE 3 | (Continued)

2. SA Prediction	
2.3 Other vehicle/road user predicted dynamic actions	**2.4 Impending hazards**
Description of other vehicle/road user E.g., one/a number of vehicles/traffic E.g., model/make/color ***Other driver/road user future maneuvers*** E.g., Overtake/turn/accelerate/continue to drive/brake/stop/pull out/wait/drive on other side of the road/signal ***Other driver future orientation/trajectory*** E.g., Straight on/round the bend/carry on/turn left or right ***Other driver future speed*** E.g., mph/faster/slow down/slow down/maintain speed ***Position relative to driver*** E.g., to "my" left/to "my" right/in front/behind/close ***Location of other vehicle/road user*** E.g., on the verge/side of the road/other side of the road ***Purpose of other vehicle/road user*** E.g., industry/occupation	***Driver general caution/awareness*** E.g., looking to see what is coming/being careful/concentrating/wary of distraction/giving wide berth when overtaking ***Specific potential hazards*** E.g., pedestrians/other road users/traffic lights/limited visibility/roundabout Future actions of other road users (e.g., walk in road, cross without looking) ***Location of projected hazard*** E.g., on the right/on the left ***Abstract/hypothetical hazards*** E.g., a crash/animals in the road ***Absence of hazards*** E.g., no bends ahead/no signs/no pedestrians/no cars present/road clear/nothing will happen ***Theory of mind projection to evaluate future risk*** E.g., being in a daydream

SA Taxonomy Derived From the Comprehension Task

The theme "1.1 *Road/highway,*" was the dominant feature in participants' descriptions, making up 30% of the total proportion of references in the dataset. Within this theme, the sub-themes "type of highway" (*n*i = 23.5%) and "road layout" (*n*i = 27.8%) contributed the most to participants' accounts of what was happening in the remote scene. All participants told us about the road that they were currently on, describing the type of road, for example a rural road, but also using adjectives such as "busy" or "fast." This theme included broad information about traffic signs from either on-road signs or signposts (*n*i = 11.7%). Remote awareness cascaded from detecting the presence of the sign ("*a sign*"), understanding what the sign meant ("*saying coming to a left-hand bend*") and perceiving the location of the sign relative to the driver ("*past the slow sign on the road*"). This indicates an interplay between the "perception" and "comprehension" SA levels which is discussed in more detail in section "Interplay Between Different Levels of SA." The nature of the locale in the driving video was also commented upon by all participants in at least one of the videos (e.g., whether it was a village or residential area), with some participants even commenting on trees and road "decoration," thus building up a vivid narrative of the environment that the car is currently occupying, a crucial requirement for navigation. Although, attention to tangential details such as these might be an example of participant variability in SA which we discuss further in section "Individual Differences in Participant SA," we suggest that these perceptual encoding details allow participants to construct a narrative of the scene assisting their comprehension.

In the theme "1.2 *Driver's perspective,*" the driver's judgment of their own action/s, whether temporal judgments of whether to pass parked cars or wait, was illustrative of the capacity for video relay to transpose space and time allowing the viewer to imagine themselves as the driver in the scene. Although we had not tasked participants to pretend to be an RO, they had been instructed to consider themselves as the "driver" and most consistently referred to "their" relative position to other cars (*n*i = 41.9%) and their current trajectory (*n*i = 26.3%). This suggests that video feed can be a powerful medium when distributing information to ROs. The detail in which participants described "their" perspective suggests that they had a good sense of presence in the scene (although we cannot know how immersive they found watching driving videos) and participants often referenced the speed that they were traveling at (*n*i = 5.6%) on busier road type videos.

The theme "1.3 *Other drivers/vehicles*" demonstrated that being aware of other drivers on the road is an important feature of driving SA as the total proportion of references to what other drivers and vehicles were doing was identical to their judgment of their own action/s in theme "1.2 *Driver's perspective*" (*n*i = 21% for both). In theme 1.3, participants most commonly mentioned descriptive details (*n*i = 55.2%) of the other driver/vehicle such as make, model and color. We also observed detailed analysis of the presence and location ("*up on the verge*") but also the state ("*abandoned*") and purpose ("*some sort of waste truck*"/" *local authority vehicle*") of other vehicles. Furthermore, participants attended to the identifying characteristics of other vehicles depicted by logos or branding on the side of the vehicle, for example in the video, Residential #2, the "driver" is following a DPD van which all participants alluded to in their verbal accounts. These data were used to conclude that it was likely that the driver would make frequent stops as they were "*looking for somewhere to park up to deliver a parcel,*" showing that theory of mind analysis (*n*i = 3%) of other road users' ongoing behavior is a component of building driving SA.

TABLE 4 | Total proportion of references to each theme and the frequency (%) at which each of the sub-themes were referenced within each theme.

SA Comprehension Macro and micro category list									Total proportion of references
1.1 Road									30%
Type of highway	**Road description**	**Type of area**	**Road layout**	**Road layout (adjective)**	**Time of day/season**	**Traffic signs**	**Road "decoration"**		
23.5%	19.6%	5.2%	27.8%	1.7%	0.9%	11.7%	9.6%		
1.2 Driver's perspective									21%
Relative position to other cars	**Direction/trajectory of driver**	**Speed traveling at**	**Subjective judgment of speed**	**Driver's decision/judgment of own action/s**	**Rationale of judgment of own action/s**	**Driver perception issues**			
41.9%	26.3%	5.6%	3.1%	7.5%	12.5%	3.1%			
1.3 Other drivers/vehicles									21%
Description	**Speed traveling at**	**Direction/trajectory**	**Purpose of vehicle**	**Location of vehicle**	**State of vehicle**	**Rationale/analysis of other driver action (theory of mind)**			
55.2%	4.2%	10.9%	6.1%	12.1%	8.5%	3.0%			
1.4 Dynamic actions of driver and/or other drivers/vehicles									23%
Maneuver planning		**Active maneuvers**		**Observation**					
Driver	**Other driver/vehicle**	**Driver**	**Other driver/vehicle**	**Driver**	**Other driver/vehicle**				
13.7%	0.6%	66.9%	16.0%	1.7%	1.1%				
1.5 Pedestrians/other road users									5%
Presence of pedestrian/other road user	**Location of pedestrian/other road user**	**Location of pedestrian/other road user relative to driver**	**Description of pedestrian/other road user**	**Gender of pedestrian/other road user**	**Age of pedestrian/other road user**	**Action of pedestrian/other road user (what pedestrian is doing)**	**Purpose of action of pedestrian/other road user (why they are doing it)**	**Occupation of pedestrian/other road user**	
22.9%	5.7%	20.0%	17.1%	11.4%	2.9%	17.1%	2.9%	2.9%	

(Continued)

TABLE 4 | (Continued)

SA Comprehension Macro and micro category list							**Total proportion of references**
1.6 Anticipatory Hazards							1%
Keeping distance from other vehicles	**Watching out for other vehicles' behavior**						
75.0%	25.0%						
SA Prediction Macro and micro category list							**Total proportion of references**
2.1 Changing road environment							13%
Road layout perception	Environmental markers	Changing road details	Traffic signs ahead				
40%	9%	36%	15%				
2.2 Driver future action/s							43%
Driver future observation	Driver future maneuvers	Driver future orientation/trajectory	Driver future speed	Driver position on the road	Reaching destination		
4%	49%	21%	12%	9%	6%		
2.3 Other vehicle/road user predicted actions							25%
Other driver/road user future maneuvers	Other driver/road user future trajectory	Other driver future speed	Position relative to driver	Location of other vehicle/road user	Description of other vehicle/road user	Purpose of other vehicle/road user	
15%	1%	10%	26%	6%	40%	1%	
2.4 Impending hazards							19%
Driver general caution/awareness	Specific potential hazards	Location of projected hazard	Abstract/hypothetical hazards	Absence of hazards	Theory of mind projection to evaluate future risk		
23%	41%	4%	15%	15%	2%		

TABLE 5 | Examples of each theme with illustrative quotes from participant transcripts.

Illustrative quotes derived from the SA Comprehension question categorized into sub-themes of the taxonomy	
1.1 Road/highway	
Traffic signs and signals	*“there was a house on the right-hand side.coming... a sign saying coming... to a left hand bend. Also there was a DPD red van in front of me... past the slow sign on the road, then we came to a built up area with houses so... we slowed down to 30...”* P6 RES#2 WITH
Road “decoration”	*“I was driving along a single lane road, going through countryside.... there were trees on either side of the road and some fences”* P1 A#2 WITHOUT
1.2 Driver’s perspective	
Relative position to other cars	*“There were cars to my right but none in the middle lane and there was a lorry ahead in the left lane, where I was but it was quite ahead*.” P2 M#1 WITHOUT
Subjective judgment of speed	*“This was a country road with trees on both sides... I believe we were going a lot faster than 30 mile per hour.”* P6 A#2 WITH
Driver’s decision/judgment of own action	*“So I am sitting there waiting. for the cars to pass on the other side of the road... waiting for a gap... for when I can pull out.” P5 RES#1 WITH*
1.3. Other drivers/vehicles	
Location and state of other vehicles	*“and was passing... what looked like an abandoned... van on the other side of the road*.” P8 R#2 WITHOUT
Rational/analysis of other driver action (theory of mind)	*“I think he is trying to find out where to drop his parcel off...”* P5 RES#2 WITHOUT
Purpose of other vehicles	*“some sort of waste truck...”* P9 R#2 WITHOUT
1.4 Dynamic actions of driver and/or other drivers/vehicles	
Active maneuvers (driver)	*“I waited my turn and then proceeded to indicate and go round the vehicle.”* P3 RES#1 WITH
Active maneuvers (other driver/vehicle)	*"the black van in front of me had stopped. . .*" P1 A1#WITH
1.5 Pedestrians/other road users	
Location of the pedestrian/other road user relative to driver	*“. . .there was a pedestrian walking on the left and there were cars approaching from the right...”* P2 RES#1 WITHOUT
Gender of the pedestrian	*“A lady with a blue shopping bag walked past at the same time.”* P3 RES#1WITH
Action of pedestrian/other road user	*“at one point a local authority worker was in an red. . . orange high vis on the side of the road clearing debris.”* P3 R#1 WITHOUT
1.6 Anticipatory Hazards	
Keeping distance from other vehicles	*“Because I’m on a bend and I had slowed down for the vehicle in front...”* P 2 Res#2 WITH
Illustrative quotes derived from the SA Prediction question categorized into sub-themes of the taxonomy	
2.1 Changing road environment	
Road layout perception	*“To go on the bend I will be slowing down my speed and being careful to make sure that... be sure because I don’t have much visibility going around the bend.” P2 R#1WITHOUT*
Environmental markers	*“. . .the vehicle will slow down to take the hump back bridge over the railway to the right... for... coming in to what I imagine is a small village or town” P3 R#1WITHOUT.*
2.2 Driver future dynamic action/s	
Driver future maneuvers	*“I will carry on overtaking the van and then pull back into my lane on the left hand side of the road. . .” P1 RES#2 WITH*
Driver position on the road	*“I imagine that I will continue down the road in the middle lane proceeding past the... vehicles on the left, returning to the outer left hand lane...” P3 M#2 WITHOUT.*
2.3 Other vehicle/road users predicted dynamic actions	
Other driver/road user future speed	*“I am coming up behind a van that is going slower than me, so I will have to brake as I come up close behind the van.” P1 A#2 WITHOUT*
Other driver/road user future maneuvers	*“... the Volkswagen golf behind me appeared to be quite close to the vehicle and may attempt to overtake” P3 A#1 WITH*

(Continued)

TABLE 5 | (Continued)

Illustrative quotes derived from the SA Comprehension question categorized into sub-themes of the taxonomy	
2.4 Impending hazards	
Driver general caution/awareness	*"I am about to drive over a hill... so I predict that I may well encounter... some form of hazard as I... go around the bend." P7 R#1 WITHOUT*
Absence of hazards	*"I think it will have been just a normal road with traffic on my right hand side... type of a country road... certainly going faster than 30 mile per hour, looking at the video. So, I actually don't think anything will happen." P6 R#2 WITHOUT*
Theory of mind projection to evaluate future risk	*""I'm hoping that she doesn't step out into the road... so I have to brake suddenly but you never know with people when they are in a bit of a dolly day dream, they might do that..." P8* Res#1 WITHOUT.

n.b. Notation details as follows: P refers to participant number, each video type is referenced (A#1, A#2, M#1, M#2, RES#1, RES#2, R#1, R#2), WITH/WITHOUT refers to whether the participant viewed the video with rear-view image present, or without rear-view image.

Both the driver and other drivers' driving operations featured heavily in the qualitative descriptions of what was happening in the driving videos, highlighted in the theme 1.4 *Dynamic actions of driver and/or other drivers/vehicles.*" This theme was easily separated into stages, where "drivers" were planning maneuvers (*n*i = 13.7%), carrying them out (*n*i = 66.9%) or making observational checks (*n*i = 1.7%). Participants' nuanced awareness of the dynamic driving process was evidently a key feature of their attentional focus, even from a remote viewing position and, in less detail, they also commented on the dynamic actions of other drivers.

We also found that consistent features of information such as presence, location and characteristics were reported in the theme "1.5 *Pedestrians/other road users,*" in a similar style to how participants told us about both road signs and other vehicles in other themes. Participants noted the *presence* of a pedestrian (*n*i = 22.9%) if there was one in the video, as well as reporting *where* they were within the remote scene relative to the driver themselves (*n*i = 20%). Participants frequently provided a physical description (*n*i = 17.1%) and what activity they were engaged in (*n*i = 17.1%) and the gender of the pedestrian (*n*i = 11.4%).

Hazard perception has been identified in SAGAT measures as belonging to Projection thus restricted to responses concerning the prediction of future events, yet we found evidence that the theme "Anticipatory hazards" firmly belonged in the SA comprehension taxonomy section. Recent analysis of freeze and probe methods such as SAGAT has suggested that the debate surrounding SA is mainly concerned with how to measure SA effectively, instead it should be focused on whether the levels can truly be considered separate (de Winter et al., 2019). Although this theme represented only 1% of the total dataset, it was evident that participants had not rigidly stuck to the "comprehension/prediction" distinction including hazards such as keeping distances from other vehicles (*n*i = 75%) in their responses to the comprehension question, suggesting that conceptualizing driving SA into three separate levels as suggested by Endsley may be an oversimplification. We discuss this further in section "Interplay Between Different Levels of SA."

SA Taxonomy Derived From the SA Prediction Task

Some participants were not confident at making predictions or found the question facile. Other participants took a far more calculated approach to assessing the situation that had occurred at the end of the video and made predictions, which logically continue from the last action/s or were derived from micro cues, such as where someone's head is turned (not just whether they are signaling in that direction). There were strikingly different strategies to predicting – carefully judging from current information how the scene can play out in the future or being prepared for any eventuality. It is impossible to say which is the better approach, as being too sure of what will unfold in the next few seconds may make you "blind" to an immediate hazard that presents itself without warning. Participant transcripts revealed that predictions are constructed from expectancies and experience, other drivers' actions, knowledge of the rules of the road and the likely behavior of other drivers.

Furthermore, in the theme "*2.1 Changing road environment*" all participants demonstrated an evaluation of whether the road would stay the same or change in the future (*n*i = 36%) – this is clearly an important consideration that also shapes the assessment of risk and future maneuvers. For example, if you are going to go around a bend you will expect to slow down and be conscious of possible hidden obstructions. This type of driving is sometimes known as defensive driving, which draws on anticipatory responses to current driving states rather than relying on the situational model (Walker et al., 2009).

The theme, "*2.2 Driver future dynamic action/s,*" dominated the total proportion of references in the data set (43%). Finishing maneuvers that had been started when the video ended, such as pulling back into the lane on the driver's side of the road were regularly cited (*n*i = 49%). In some of the road types such as the motorway videos, participants also told us about future positioning on the road, for example which lane they would be occupying. We also saw temporal references whereby participants were analyzing the spatial distances that they were physically able to traverse based on their remote estimation of speed. For example, Participant 4 decided that there was enough time left on the "green" light to get through. If they were remotely controlling the vehicle, they may have put their foot on the accelerator to ensure this happened, but they may have been mistaken. This would be an important consideration for a RO. If their temporal judgment were negatively affected by the remote location, this could have dangerous consequences.

In the theme, "*2.3 Other vehicle/road user predicted dynamic actions,*" participants framed their future actions in the context of other road users, revealing an awareness that their future maneuvers did not happen in isolation. Description of the other road user (*n*i = 40%) and their relative position to the driver (*n*i = 26%) were the main subjects of this theme showing perceptual and spatial awareness of the total road environment was being drawn to make predictions about what would happen next.

Participants mentioned specific potential hazards (*n*i = 41%) far more frequently in the SA Prediction question than in the SA Comprehension question which was recorded in the theme, "*2.4 Impending Hazards*" although this theme contributed only 19% to the total dataset. An ambiguous sense of "being careful" was a regular consideration shown by all participants in response to the SA Prediction question. Interestingly, the *absence* of stimuli was also noticed and commented on (*n*i = 15%) as often as hypothetical hazards (*n*i = 15%), for example if there is no sign warning for a bend, we can assume that it is a straight road ahead thereby eliminating consideration of the potential hazard of a sharp bend.

Qualitative Findings Relating to SA in Remote Driving Contexts

In addition to developing the taxonomy, we were also able to interrogate participant transcripts to identify patterns that were apparent in the construction of remote SA of driving videos. The following section focuses on the overarching patterns that surfaced during the analysis of participant transcripts which illustrate the complexity of building SA in remote scenes.

SA of Participants in Relation to Absence or Presence of Rear-View Information

One aspect of our investigation was to compare SA of participants in the presence and absence of rear-view information. Some participants appeared to incorporate the rear-view information into their SA, either through indirect references to perceptual information "behind" them or to predict what the car following them may do next. However, no clear effects of this manipulation were evident in the transcripts and qualitative comments in the Debrief showed disagreement about whether a rear-view mirror was useful or was ignored completely (shown in **Table 6**).

Individual Differences in Participant SA

There were noticeable differences between participants in what they reported from the driving videos even though they all saw the same scenes. Some participants were descriptive and some extremely succinct, for example, Participant 7 reported 90% less than Participant 5. To illustrate this point, in the video Residential #1, some participants merely described the presence and location of the pedestrian, whereas others gave additional details about her gender (*"lady*"), age ("*young*") and what she is carrying ("*shopping bag*"). This could be evidence of individual speaking style, whereby some people use enhanced linguistic codes to provide "*verbal elaboration of meaning*" so some of the differences may relate more to speaking style than actual underlying awareness (Bernstein, 1964, p. 630). Other discrepancies in the verbal reports relate to differences in underlying awareness and subjective judgments as what is important in the scene. For example, a salient feature of video R#2 is an abandoned refuse truck parked on the verge by the side of the road (as the driver may emerge later down the road which could present a hazard) but a minority of participants neglected to mention its presence at all.

There were also noticeable differences between what participants observed in the driving videos. Some participants focused almost exclusively on road features such as road layouts and trees, whereas others reported speed limits very consistently. These differences between participants' attention and reporting styles appeared to remain fairly consistent within participants – someone who observes types, makes and colors of vehicles in one road makes similar observations on all the videos shown, whereas someone else never alludes to these details at all, but instead shows consistent attention to the location or purpose of other vehicles. For example, only three participants referenced details of "road decoration," a sub-theme in "1.1 *Road/highway,*" but they did so consistently in every video which contained houses or trees. This demonstrates potential participant variability in building SA in remote contexts.

Understanding of Remote Vehicle State

Participants made frequent subjective estimations to the speed that "they" were traveling at, which tells us even through second-hand, indirect cues (the monitor view) they had a sense of traveling at speed.

> *"This was a country road with trees on both sides. I believe we were going a lot faster than 30 mile per hour." P6 A#2 WITH*

However, in some cases (including this example) their beliefs were incorrect, as the videos were filmed at the exact speed limits (or less dependent on traffic conditions) of each road.

Awareness of the spatial limitations of the car was also evident. For example, in video Residential #2, the "driver" is waiting behind a queue of parked cars until there is a gap in traffic, as the road is not wide enough for two cars to pass, and all participants demonstrated understanding of this restriction.

> *"I have to come to a stop because there are cars parked on my side of the road. and that means that there is only one side of traffic that can get along the road. So I am sitting there waiting. for the cars to pass on the other side of the road. waiting for a gap. for when I can pull out." P5 RES#1 WITH*

Although higher levels of precision would be required in order for an RO to maneuver the vehicle safely in contexts such as these, the general principles of driving and knowledge of road size may be employed to make practical SA decisions in remote contexts relating to three-dimensional navigation.

TABLE 6 | Attitudes to the presence or absence of the rear-view mirror in the videos as indicated in the debrief.

Participant comments in the Debrief concerning the rear-view mirror	
Participant 6	*"I liked the rear-view camera as people should be using it more. I found that I still used it and also can keep eyes on the road for more fullness."*
Participant 1	*"When the rear-view image appeared the first time it confused me as I didn't know what it was so I ignored it. When it appeared the second time I paid more attention and realized what it was and I definitely gave a more detailed description and remembered more."*
Participant 9	*"I don't think I looked at the rear-view camera as I wanted to concentrate on the road ahead. There was only one point that I recall looking at the rear-view camera and that was when "I" was stopped behind a parked car to see how many vehicles were stopped behind me (and drivers potentially getting grumpy if I took too long to move)."*

Theory of Mind Analysis in SA Comprehension and SA Prediction

Being aware of other drivers on the road is an important feature of SA and safe driving. Participants allocated the same proportion of the time to reporting what other drivers and vehicles were doing as they did to describing their own maneuvers. Being able to predict other people's behavior by explaining their actions as a product of their independent mental state is known as having a theory of mind (Gallagher and Frith, 2003). Participants demonstrated theory of mind (TOM) in both SA Comprehension (1.3 Other drivers/vehicles: Rationale/analysis of other driver action') and SA Prediction (2.4 Impending hazards: Theory of mind projection to evaluate future risk) when trying to analyze the reasons for other drivers their behavior (see **Table 5** for quotes in each of these themes).

Although the frequency of these references was low (only 3% in 1.3 Other driver's/vehicles category and 2% in 2.4 Impending hazards) this sub-theme is highly relevant to interpreting how RO SA is constructed. An important theoretical consideration in thematic analysis is that crucial themes may have few occurrences, yet contribute toward a greater understanding of the behavior or phenomenon (Braun and Clarke, 2013). Although rare, participants were seeking a rationale/analysis of other drivers' actions, such as why the driver is slowing down (e.g., looking for an address) or parking (e.g., driver is going to deliver a package). TOM analyses were also important in relation to pedestrians, for example when considering whether they were likely to step into road without looking. The importance of interpreting the intentions of other road users for achieving effective SA cannot be underestimated in remote contexts.

Role of Context in Comprehension Illustrated by Use of Schemas

Endsley (2000) included schema as an integral part of the mental model triggered to develop comprehension and prediction of the scene even though misapplying information or filling in missing information can lead to mistakes or prediction errors. For example, in the video A #2, one participant commented in the SA Prediction transcript that, because she saw an ambulance earlier in the video, she was expecting to see an accident on the road ahead which did not materialize.

Participants in our study also made use of context to expand their comprehension of what was happening in the driving scene drawing on generalized "schemas" of driving to describe actions that they could not have "known" they were doing from the video, such as checking mirrors and signaling:

> *". . . I was just stuck there for a long period of time waiting to indicate to go round some vehicles parked on the left hand lane. A lady with a blue shopping bag walked past at the same time. I waited my turn and then proceeded to indicate and go round the vehicle." P3 RES#1 WITH*

Deriving context may enhance comprehension when viewing videos of driving scenes as it can be used to make predictions about the likely actions of other vehicles or road users. Although the type of area in 1.1 Road/highway was only mentioned 5.2% of the time in the total dataset, for the video Residential #1 it was described by all participants. This may be because this detail provides information about the likelihood of pedestrians being present, that traffic may be more congested, even that schools may be in the locale. These same mental models would be unlikely to be produced in rural areas so, comparatively, this experiential data may feed into the SA process in some contexts more than others.

Parallel Processing in Building SA of Video Driving Scenes

Our analysis uncovered evidence that a complex range of feed-forward and feedback processing is engaged to acquire SA of video driving scenes. We can see this process in the following excerpt,

> *"I'm hoping that she doesn't step out into the road. so I have to brake suddenly but you never know with people when they are in a bit of a dolly day dream, they might do that. so I will continue along the road not only looking at cars parked in front of me and cars coming the other way so there is only room for one car to pass the parked car. but also keeping an eye on this young lady until I have passed her."* Participant 8

Awareness of the pedestrian (SA Perception) is used to predict expected outcomes ("*step out into the road*") (SA Prediction), incorporating theory of mind ("*when they are in a bit of a dolly daydream*") (SA Comprehension) and drawing on gender stereotyping and other schema-based reasoning. At the same time, the participant is constructing a narrative of a likely future ("*keeping an eye on this young lady*") which also draws on further perceptual processing (age of the pedestrian "*young*") which again may be used to estimate the likelihood of the event occurring (SA Comprehension/SA Prediction). This implies that SA in driving involves parallel processing, whereby SA Comprehension and SA Prediction can be developed simultaneously, rather than the serial progression through the levels that is suggested by previous models.

Interplay Between Different Levels of SA

SA has, thus far in Psychology, been regarded as the *total amount* of information in our possession about our environment, but how these information points relate to each other is just as important, as an interaction between the different levels of SA occurs to develop the picture of the scene (Walker et al., 2009). Endsley (2017b) maintains that the levels of SA are ascending, for example, perception feeds in to comprehension, however, we found evidence that a "higher" level can feed downward to a "lower" level. SA Comprehension or SA Prediction may influence perception; "drivers" may perceive something because of a contextual detail that they had comprehended earlier. For example, participants may be making sense of the highway layout,

> "*A single carriage road. speed limit. I think was 50 miles per hour. so although there were single lanes. there were quite a lot of. filter lanes for turning to the right.*"

But then recognize the filter lanes as they pass them, demonstrating SA perception being fed by comprehension,

> *" so I had just passed two of those filter lanes to turn to the right."*
> Participant 9

Here perception is seamlessly integrated into comprehension in both a feed-forward and feedback association. This exemplifies the clear interplay between perception and comprehension SA in driving.

In driving, estimating and analyzing hazards is important, particularly in relation to identifying driver adjustments that are necessary to prevent potential hazards from developing into actual hazards (for example leaving adequate stopping distances or slowing down and looking for oncoming traffic). Hazard perception has been identified in SAGAT measures as belonging to Level 3 Projection (see also Horswill and McKenna, 2004) and therefore restricted to responses concerning the prediction of future events. Our research instead suggests that elements of prediction are embedded in building comprehension of a remote scene. This can be seen in the example below,

> *"I am on a bend turning right and there are cars passing me on the right. To go on the bend, I will be slowing down my speed and being careful to make sure that. be sure because I don't have much visibility going around the bend."* P 2 ("Rural #1").

Although there may have been a perceptual trigger in the immediate environment to prompt an expectation that something may develop in the future (such as a road bend), the projected hazard (lack of visibility) is not yet immediately located in the scene, which challenges the temporal nature of Endsley's model, which states that SA is

> *"the perception of the elements of the environment within a volume of space and time, the comprehension of their meaning, and the projection of their status in the near future"* (Endsley, 1988, p. 792).

Instead, an emerging scene in the "now" ties to "future" predictions through a feedback and feed forward process. In summary, separating driving SA into entirely distinct levels may oversimplify the nuanced understanding of the remote scene required to build up, maintain and update SA when drawing information from remote contexts.

DISCUSSION

Comparison With Previous Research

The current study took a novel approach to measuring SA in remote driving, employing a qualitative verbal elicitation task to uncover, in its most basic form, what people "see" in a remote scene when they are not constrained by rigid questioning. Participants provided an abundance of information and qualitative detail enabling the construction of a taxonomy encompassing the types of information that are typically derived from a remote scene.

Previous models of SA have suggested that experts may have better ability than novices in building effective SA (for example hazard perception has been shown to be faster for experienced drivers compared to novice drivers (Huestegge et al., 2010; Gugliotta et al., 2017)), but few have investigated the variability between individuals in *how* they derive situation awareness in a scene, what information they sample and how much importance they give to different pieces of information in the environment. In this study, we observed differences between participants whereby, for example, some focused almost exclusively on road features (e.g., road layout, trees) whereas others reported speed limits very consistently. These patterns of attention and reporting style appeared to be consistent *within* participants – someone who observed types, makes and colors of vehicles in one video tended to make similar observations on all the videos shown, whereas someone else may never allude to these details at all, but instead direct attention to the location or purpose of other vehicles for each video.

No clear effects of the manipulation between presence and absence of rear-view information were manifest in the transcripts. Studies which have used eye tracking to measure the attention participants have given to the road ahead and the mirrors whilst viewing driving videos report a decrease in glance frequency to the rear-view mirror over time (Unema et al., 2005; Over et al., 2007; Lu et al., 2017). It is possible that participants had been accessing the rear-view information during the video but had not used the rear-view mirror recently. If they did not volunteer any information relating to that view in response to either SA Comprehension or SA Prediction questions, even though they had looked at it during the video, the qualitative measures used in this study would be unable to uncover this behavior. We may have found that participants struggled more in the absence of the rear-view mirror if they had been required to do anything active in the scene rather than passively describing it.

We note that, in addition to the provision of a rear-view feed, designers of interfaces for ROs of AVs are likely to end up providing a sophisticated data overlay including multiple sources of information about the car and the road environment to assist in the creation and maintenance of SA. Given that this is such a new area of research, we focused here on the role of video relays in developing awareness of a remote scene. Current research in our laboratory is investigating the role of both rear view and auditory feeds in supporting the development of remote SA (see Mutzenich et al., 2021, in preparation) and future work to investigate the impacts of additional information overlays will also be essential.

Endsley's original approach implied a clear separation between the three SA levels and has been criticized by many researchers for this distinction (see Sorensen et al., 2011; Salmon et al., 2012). More recently, Endsley has dismissed this characterization as a "fallacy" and acknowledges a high level of integration between the different levels, arguing that the three levels are "*ascending*"; you can go back and search for perceptual data to back up comprehension and projection (Endsley, 2015, p. 8). However, the quantitative nature of SAGAT queries forces a starker distinction between the levels in order to enable researchers to score them separately, implying that perception can be assessed separately from comprehension and so on.

The qualitative methodology used in this study allowed a freer and less constrained investigation than is possible using the quantitative SAGAT-based metrics. For example, although the transcripts were analyzed separately for SA Comprehension and SA Prediction questions, participants typically appeared to blend this information organically; for example, awareness of "hazards" was apparent in both questions, suggestive of the fact that that building up SA in driving contexts is a flexible and fluctuating process of combining comprehension and prediction globally rather than serially, as has sometimes been implied by previous SAGAT methodologies (Jones and Endsley, 1996; Endsley, 2000, 2017b). In addition, the individual variability in reporting detail of the scene is likely to be missed by SAGAT paradigms. To our knowledge, no other research to date has directly measured how SA is freely constructed in driving contexts.

Additionally, a novel and positive aspect of the naturalistic qualitative analysis used was the ability to extract unexpected information about what participants see (or don't see) in remote driving scenes, which is not possible using approaches in which participants are constrained by questions. The sub-theme "*Absence of hazards*" in the SA Prediction theme "2.4 *Impending Hazards*" illustrates that anyone reporting "no risk" would receive a score of zero on a SAGAT test because no specific items or events are mentioned, yet this is exactly the type of careful data gathering from the remote scene that would be desirable in an RO and would constitute accurate prediction. Similarly, the sub-theme, "*Subjective judgment of speed* e.g., fast/too fast/slowly" in the Comprehension theme "1.2 *Driver's perspective*" would also be missed by SAGAT as although queries may ask participants to report actual speed (mph) there is no insight into whether the participant has judged that speed to be appropriate. The taxonomy illustrates myriad observations, calculations and adjustments required in driving scenarios which have gone unnoticed in other research paradigms. This type of thorough understanding of what SA comprises in remote contexts is essential before we can start to make judgments of what "good" SA may look like and assess whether quantitative metrics can accurately judge whether or not someone has good SA.

Limitations of Research

There may have been issues related to the sample of participants used in our study. We did not issue a test on Highway Code to participants prior to the study, so there may have been a variable base level of knowledge between them. However, there were no clear individual or subgroup differences apparent which may have influenced what people spoke about, thus influencing the data that fed into the taxonomy and there was no evidence to indicate that their performance would not be mirrored in another group. In addition, although the sample size of participants ($n = 10$) was small, the total transcripts in response to the video prompt questions amounted to over 8,000 words which presented a rich and diverse dataset.

Another potential limitation concerns the possibility that the use of an example video depicting the "rear absent" condition at the start of the study may have discouraged participants from providing rear view information in response to future videos where this information was present. This seems unlikely because all participants mentioned information at some point that was present in the rear-view feed.

A further consideration is that in this task participants were not actively driving and so were not subject to attention being diverted by other driving tasks such as changing gears, depressing pedals and other peripheral distractions (e.g., changing music). Viewing remote video driving scenes may add a sense of speed or at the very least, distort the viewers' sense of motion. Without force feedback pushing you back into the seat or information from the tire friction on the road, it is difficult to accurately judge how fast you are driving and to remain engaged in the driving task (Tang et al., 2013). Although the results cannot represent authentic driving experiences, ROs are unlikely to be carrying out these tasks either, so from that perspective this study may not differ greatly from the task of an RO. Mackenzie and Harris (2015) showed that participants were slower to detect hazards when driving themselves as opposed to passively viewing a video, so RO SA may be augmented by the detached nature of their location and the singular nature of their task – to work out from the driver's perspective what the problem is, via second hand information from the scene.

Some themes in the taxonomy were less prevalent than others, but this may have reflected the specifics of the videos used, rather than the overall nature of the participants' SA itself. For example, theme "1.5 *Pedestrians/other road users*" represented only 5% of the total data set but this could be because very few of the videos contained images of pedestrians and other road users, as the videos were predominantly filmed on busy highways and motorways, yet all participants mentioned pedestrians when they were present in the video. In addition, the taxonomy did not include any sub-themes relating to roadworks which would be highly relevant to remote operators (as outlined in the introduction to this paper, common challenges to AVs are deviations from their path required by construction works which may be signaled by workers using hand gestures). The naturalistic element of the stimuli used in this study meant that no roadworks were encountered while the videos were being filmed. Future stimuli should incorporate a wider range of driving scenarios, such as busy urban streets, to enable the further generalization of the themes identified. There is also scope for future research to test some of the smaller sub-themes to see if they are replicated by future observers.

For example, the sub-theme *"Theory of mind"* in both SA Comprehension and SA Prediction taxonomies did not feature in many participants' responses yet presented an important finding in relation to how SA is constructed in a "sense making" context. In thematic analysis methodology, researchers must use their judgment as to how many occurrences of an item are necessary for it to be important (Braun and Clarke, 2013). Indeed, we propose that TOM is a critical element involved in processing and comprehension of a driving scene. Spiers and Maguire (2007, p. 1675) also found examples of TOM considerations (which they termed "spontaneous mentalizing," referring to participants thinking about the thoughts of others) in their study of London taxi drivers engaging in a verbal description protocol whilst watching/playing a realistic first-person driving game in an MRI scanner. SA Comprehension may require frequent transitions from the driver's perspective to imagining another driver's intentions, then acting on that information. Thus, it will also be necessary for ROs to practice TOM if required to engage dynamically with the remote driving task. Future stimuli should include more direct interactions with other road users and other vehicles to ascertain whether this sub-theme would have more prevalence in these driving situations and give more support to its inclusion in the taxonomy.

As the study was hosted online, it was necessary to present the videos in a low resolution (811 × 456) to enable them to be cached and streamed using the hosting site Gorilla.sc. Although ROs are likely to receive higher resolution video and on a larger format than desktops or laptops there was no evidence from participants' performance feedback in the debrief to suggest that the quality of the display was insufficient for them to extract the necessary information. However, it would be beneficial in future studies to set a standard screen size to give a broader and more HD viewing experience which may make it more immersive or make use of VR to present the visual information as if the driver were *in situ*.

The use of inductive thematic analysis methodology involved an element of subjectivity when classifying and grouping themes. Future research may wish to extend this work with potentially more objective analyses. For example, it has been suggested that network analysis can be used to model SA, by identifying first the information underpinning SA such as "noun-like information elements" and then establishing relationships between different pieces of information (Walker et al., 2009, p. 680; Salmon et al., 2013).

Conclusion and Recommendations for Future Work

There are several projects currently underway in the United Kingdom designed to understand a range of aspects of remote operation, including human factors considerations, with the aim of enabling remote operation to become a feasible transport opportunity. This research contributes toward the knowledge that will enable the acceptance of autonomous technology in the future.

More specifically, adopting a more nuanced approach to considerations of situation awareness could improve the design of remote operation support systems. Our research suggests that designers of interfaces for ROs should take careful consideration of the scope and range of the information derived from the remote driving videos and also the variability between participants in how they construct situation awareness of remote scenes. We recommend engaging in iterative research processes before and after implementing new graphical user interfaces to ensure that ROs are given information that has been empirically proven to be useful in their SA development. This knowledge may also contribute to the further development of industry standard taxonomies for remote operation so that regulatory frameworks can be established with regards to the training and technology necessary to carry out the role.

Another interesting use of the taxonomy would be to determine which sub-themes are indispensable in remote SA and which are "nice to have," from the perspective of the actions of remote operators. For example, in the theme "1.3 Other vehicles/road users," having awareness of the make, model, color etc., of vehicles on the road around you may not be as crucial as understanding the location or trajectory of the vehicles around you on the road. Yet, if there was a crash incident, details such as these would be important for the identification and reporting of other vehicles at the scene. Determining which are the minimum requirements for remote operator SA will be important for selection and training of ROs in the future.

The evidence in this study provides a rich catalog of verbal data that exemplifies the interactions between different SA levels that operate when participants process information from a remote naturalistic driving scene. Our open-source videos of remote driving situations can play a role in developing further understanding of the unique SA requirements for ROs, supporting the construction of new SA questions based on the information that participants in this study have been shown to extract from the videos. The proposed taxonomy can also be used in future empirical work to design queries that can effectively measure RO SA. This work is currently underway in our lab.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Royal Holloway, University of London Psychology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CM wrote the manuscript. PD, SD, and SH reviewed several drafts and contributed to revisions of the manuscript. All authors approved the final manuscript.

FUNDING

Funding source—ESRC Doctoral Training Partnership with South East Network for Social Sciences (SeNSS). Grant Reference Number: ES/P00072X/1.

ACKNOWLEDGMENTS

Thanks to the three reviewers who made helpful and insightful suggestions to improve the manuscript.

REFERENCES

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. (2018). Gorilla in our Midst: An online behavioral experiment builder. *Behav. Res.* 52, 388–407.

Banks, V. A., Eriksson, A., Donoghue, J. O., and Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Appl. Ergonom.* 68, 138–145. doi: 10.1016/j.apergo.2017.11.010

Bernstein, B. (1964). Elaborated and Restricted Codes: Their social origins and some consequences. *Ethnogr. Commun.* 66, 55–69. doi: 10.1525/aa.1964.66.suppl_3.02a00030

Bolstad, C. A., Cuevas, H., Wang-Costello, J., Endsley, M. R., and Angell, L. S. (2010). "Measurement of situation awareness for automobile technologies of the future," in *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*, ed. G. L. Rupp (Warrendale: SAE International), 195–213.

Boyatzis, R. E. (1998). *Transforming qualitative information: thematic analysis and code development*. Thousand Oaks, CA: SAGE.

Braun, V., and Clarke, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *Psychologist* 26, 120–123.

Buscher, G., Cutrell, E., Morris, M. R., and Way, O. M. (2009). "What do you see when you're surfing? Using eye tracking to predict salient regions of web pages," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (New York, NY: ACM), 21–30.

Capallera, M., Meteier, Q., De Salis, E., Angelini, L., Carrino, S., Khaled, O. A., et al. (2019). *Owner manuals review and taxonomy of ADAS limitations in partially automated vehicles. Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI*. Utrecht, Netherlands: ACM.

de Winter, J. C. F., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., and Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cogn. Technol. Work* 21, 99–111. doi: 10.1007/s10111-018-0527-6

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Hum. Fact. Soc. Proc. Hum.* 32, 97–101. doi: 10.1177/154193128803200221

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Hum. Factors* 37, 65–84.

Endsley, M. R. (2000). "Theoretical underpinnings of situation awareness: A critical review," in *Situation awareness analysis*, eds M. R. Endsley and D. J. Garland (Mahwah, NJ: Lawrence Erlbaum Associates).

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *J. Cogn. Engin. Decis. Mak.* 9, 4–32. doi: 10.1177/1555343415572631

Endsley, M. R. (2017b). Toward a theory of situation awareness in dynamic systems. *Hum. Error Aviat.* 2017, 217–249. doi: 10.4324/9781315092898-13

Endsley, M. R. (2017a). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *J. Cogn. Engin. Decis. Mak.* 11:155534341769519. doi: 10.1177/1555343417695197

Endsley, M. R. (2019). Situation awareness in future autonomous vehicles: Beware of the unexpected. *Adv. Intel. Syst. Comp.* 824, 303–309. doi: 10.1007/978-3-319-96071-5_32

Endsley, M. R., Bolte, B., and Jones, D. (2003). *Designing for Situation Awareness*, 2nd Edn. London: CRC Press, doi: 10.1201/9780203485088

Franz, B., Haccius, J. P., Stelzig-krombholz, D., Pfromm, M., and Kauer, M. (2015). "Evaluation of the SAGAT method for highly automated driving," in *Proceedings of the 19th Triennaial Congress of the IEA, August, 1–8*, Melbourne, VIC.

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind.'. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6

Graneheim, U. H., Lindgren, B. M., and Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Educ. Today* 56, 29–34. doi: 10.1016/j.nedt.2017.06.002

Gugerty, L. J. (1997). Situation awareness during driving: explicit and implicit knowledge in dynamic spatial memory. *J. Exper. Psychol. Appl.* 3, 42–66. doi: 10.1037/1076-898X.3.1.42

Gugerty, L. J. (2011). "Situation awareness in driving," in *Handbook for Driving Simulation in Engineering, Medicine and Psychology*, eds J. Lee, M. Rizzo, D. Fisher, and J. Caird (Boca Raton, FL: CRC Press), 265–272.

Gugliotta, A., Ventsislavova, P., Garcia-Fernandez, P., Peña-Suarez, E., Eisman, E., Crundall, D., et al. (2017). Are situation awareness and decision-making in driving totally conscious processes? Results of a hazard prediction task. *Transport. Res. F Traffic Psychol. Behav.* 44, 168–179. doi: 10.1016/j.trf.2016.11.005

Horswill, M. S., and McKenna, F. P. (2004). "Drivers' hazard perception ability: situation awareness on the road," in *A Cognitive Approach to Situation Awareness*, eds S. Banbury, and S. Tremblay (Aldershot: Ashgate), 155–175.

Huestegge, L., Skottke, E. M., Anders, S., Müsseler, J., and Debus, G. (2010). The development of hazard perception: Dissociation of visual orientation and hazard processing. *Transport. Res. F Traffic Psychol. Behav.* 13, 1–8. doi: 10.1016/j.trf.2009.09.005

Jansen, R. G., Wiertz, L. F., Meyer, E. S., and Noldus, L. P. J. J. (2003). Reliability analysis of observational data: Problems, solutions, and software implementation. *Behav. Res. Methods Instruments Comp.* 35, 391–399. doi: 10.3758/BF03195516

Jewitt, C. (2012). *National Centre for Research Methods Working Paper: An introduction to using video for research*. London: National Centre for Research Methods.

Jones, D., and Endsley, M. R. (1996). Sources of situation awareness errors in aviaiton domain. *Aviat. Space Environ. Med.* 67, 507–512. doi: 10.1039/c4qo00187g

Khattak, A. J., Ahmad, N., Wali, B., and Dumbaugh, E. (2021). A taxonomy of driving errors and violations: Evidence from the naturalistic driving study. *Accid. Anal. Prev.* 151:105873. doi: 10.1016/j.aap.2020.105873

Larsson, A. F. L., Kircher, K., and Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transport. Res. F Traffic Psychol. Behav.* 27, 229–237. doi: 10.1016/j.trf.2014.05.008

Lo, J. C., Sehic, E., Brookhuis, K. A., and Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transport. Res. F Traffic Psychol. Behav.* 43, 325–338. doi: 10.1016/j.trf.2016.09.006

Lu, Z., Coster, X., and de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Appl. Ergon.* 60, 293–304. doi: 10.1016/j.apergo.2016.12.003

Ma, R., and Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *Int. J. Indust. Ergonom.* 35, 939–953. doi: 10.1016/j.ergon.2005.04.002

Ma, R., and Kaber, D. B. (2007). Situation awareness and driving performance in a simulated navigation task. *Ergonomics* 50, 1351–1364. doi: 10.1080/00140130701318913

Mackenzie, A. K., and Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Vis. Cogn.* 23, 736–757. doi: 10.1080/13506285.2015.1079583

Matthews, M. L., Bryant, D. J., Webb, R. D. G., and Harbluk, J. L. (2001). Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems. *Transport. Res. Rec.* 1779, 26–32. doi: 10.3141/1779-04

McHugh, L. M. (2012). Interrater reliability: the kappa statistic. *Biochem. Med.* 22, 276–282.

Mutzenich, C., Durant, S., Helman, S., and Dalton, P. (2021). Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles. *Cogn. Res. Princ. Implic.* 6:9. doi: 10.1186/s41235-021-00271-8

Niklasson, L., Riveiro, M., Johansson, F., Dahlbom, A., Falkman, G., Ziemke, T., et al. (2007). "A Unified Situation Analysis Model for Human and Machine Situation Awareness," in *Informatik 2007 – Informatik Trifft Logistik – Band 2*, eds O. Herzog, K.-H. Rödiger, M. Ronthaler, and R. Koschke (Bonn: Gesellschaft Für Informatik e. V), 105–109.

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., and Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vis. Res.* 47, 2272–2280. doi: 10.1016/j.visres.2007.05.002

Revell, K. M. A., Richardson, J., Langdon, P., Bradley, M., Politis, I., Thompson, S., et al. (2020). Breaking the cycle of frustration: applying Neisser's perceptual cycle model to drivers of semi-autonomous vehicles. *Appl. Ergon.* 85:103037. doi: 10.1016/j.apergo.2019.103037

SAE International (2016). *Automated Driving: Levels of Driving Automation are Defined in New SAE International Standard J3016*. Warrendale, PA: SAE International.

SAE International (2021). *SAE J3016; Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Warrendale, PA: SAE International.

Salmon, P. M., Stanton, N. A., and Young, K. L. (2012). Situation awareness on the road: Review, theoretical and methodological issues, and future directions. *Theor. Issues Ergonom. Sci.* 13, 472–492. doi: 10.1080/1463922X.2010.539289

Salmon, P. M., Young, K. L., and Cornelissen, M. (2013). Compatible cognition amongst road users: The compatibility of driver, motorcyclist, and cyclist situation awareness. *Safety Sci.* 56, 6–17. doi: 10.1016/j.ssci.2012.02.008

Scholtz, J. C., Antonishek, B., and Young, J. D. (2005). "Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces," in *Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, Vol. 35, (Piscataway, NY: IEEE), 450–459. doi: 10.1109/TSMCA.2005.850589

Sorensen, L. J., Stanton, N. A., and Banks, A. P. (2011). Back to SA school: Contrasting three approaches to situation awareness in the cockpit. *Theor. Issues Ergonom. Sci.* 12, 451–471. doi: 10.1080/1463922X.2010.491874

Spiers, H. J., and Maguire, E. A. (2007). Neural substrates of driving behaviour. *NeuroImage* 36, 245–255. doi: 10.1016/j.neuroimage.2007.02.032

Stanton, N. A., and Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Sci.* 47, 227–237. doi: 10.1016/j.ssci.2008.03.006

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Pract. Assess. Res. Eval.* 9, 66–78.

Störmer, V. S. (2019). Orienting spatial attention to sounds enhances visual processing. *Curr. Opin. Psychol.* 29, 193–198. doi: 10.1016/j.copsyc.2019.03.010

Tang, T., Kurkowski, J., and Lienkamp, M. (2013). Teleoperated road vehicles: A novel study on the effect of blur on speed perception. *Int. J. Adv. Robot. Syst.* 10:56735. doi: 10.5772/56735

UNECE (2020). *Human Factors Challenges of Remote Support and Control: A Position Paper from HF-IRADS. Informal Document No. 8 - September 2020 Session of WP.1*. 1–9. Available Online at: https://www.unece.org/fileadmin/DAM/trans/doc/2020/wp29grva/GRVA-07-65e.pdf

Unema, P. J., Pannasch, S., Joos, M., and Velichkovsky, B. M. (2005). Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. *Vis. Cogn.* 12, 473–494.

Walker, G. H., Stanton, N. A., and Young, M. S. (2008). Feedback and driver situation awareness (SA): A comparison of SA measures and contexts. *Transport. Res. F Traffic Psychol. Behav.* 11, 282–299. doi: 10.1016/j.trf.2008.01.003

Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., and Jenkins, D. P. (2009). Does advanced driver training improve situational awareness? *Appl. Ergonom.* 40, 678–687. doi: 10.1016/j.apergo.2008.06.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Paper 3

Development and validation of a new SA measure for remote driving videos

Abstract

Situation awareness (SA) needs during remote operation will differ to SA requirements during in-situ driving, meaning that standard measurements to assess SA in driving may not be appropriate. This paper aims to address the lack of standardisation in previous quantitative driving-related SA measures. Using the Taxonomy of Situation Awareness (TSA) developed in previous research (see Mutzenich et al, 2021), we designed SA questions for sixteen original driving videos. Each question was designed to measure a particular level of SA (perception, comprehension, and prediction). We conducted a pilot study (n = 12) to trial the new remote SA measure, using an iterative qualitative and quantitative approach to revise and refine individual SA queries. We then utilised the SA measure in an experiment (*n= 94*) and conducted an exploratory factor analysis to identify the underlying constructs of the instrument and to validate whether the questions successfully measured remote SA. Results suggest that the SA measure has four underlying constructs: factor 1 represents spatial and environmental awareness, factor 2 represents anticipatory hazards, factor 3 dynamic driving actions and factor 4 represents other road users. Internal consistency of the criterion set was evaluated by Cronbach's alpha coefficient at α= 0.832, 95% CI [0.734, 0.896]. Factor loadings demonstrated that SA questions on perception, comprehension and prediction are combined when building awareness of a remote driving scene, which suggests that measurement tools designed to hold SA levels separate may be unsuitable in driving contexts. Our factor structure also corresponded with previous categorisation of the SA concepts that underpin remote driving SA from the TSA. Our research goes towards developing a validated tool to measure the SA needs of remotely operating a vehicle.

1.0 Introduction

Autonomous driving applications are becoming increasingly common across a broad range of industries, for example in freight and logistics, agriculture, mining, and shipping. Since the start of 2021, there has been a marked increase in the UK of trials of autonomous mobility solutions, such as low-speed autonomous delivery vehicles for high street retailers (wilko though StreetDrone[2]), app-based ride-hailing services (Fetch through Imperium Drive[3]) and autonomous Stagecoach bus shuttles operating a 30-mile route over the Forth Road Bridge as part of Project CAVForth (in partnership with

---

[2] https://www.streetdrone.com/

[3] https://imperiumdrive.com/

Fusion Processing[4]). Each of these ventures, together with others like them, combine autonomous technology with a remote operator (RO), who may be required to drive the vehicle remotely if the automation fails or requires human assistance.

1.1 Measuring situation awareness in remote driving

There are significant human factors challenges unique to drivers remotely in charge of a vehicle. ROs will have to build up a mental model of the remote environment, known as situation awareness (SA) facilitated primarily by video feed. As defined by Endsley (1988), SA can be sub-divided into ascending levels of perception, comprehension and projection. SA encompasses what is known about the environment, what is happening in it and what might change in the near future. The task of developing SA in driving contexts from a remote location is likely to be made more difficult by the operator's physical absence. It is likely to take more time for an RO to gather important SA information about environmental factors such as weather and road conditions, vehicle-related factors such as speed and heading direction, and positional factors, such as relative distances to other objects compared to in-situ SA.

The question of how to effectively measure SA needs during remote operation represents an operational challenge in the autonomous transport field. This is an important endeavour, because the automated vehicle industry will need a unified approach to measuring remote SA if regulatory frameworks are to be established with regard to the necessary technology and training required to achieve adequate remote SA. The Situation Awareness Global Assessment Technique (SAGAT) is broadly recognised as an objective measure of SA and has been applied in many domains such as the army, marine and military, aviation, healthcare, and power industries (Endsley, 1988, Endsley, 2021). Its methodology involves participants taking part in simulated trials answering generic queries designed to measure all three SA levels. SA performance is measured by comparing the accuracy of their responses to the queries with the reality of the simulation.

Although originally designed to measure SA in pilots, SAGAT has been used in the driving domain to measure driving SA as it is claimed that queries can be applied generically to any situation. Endsley, Bolte, & Jones (2003) developed the quantitative SAGAT queries by conducting a Goal Directed Task Analysis (GDTA), generating a comprehensive list of the information that operators need to conduct their role in any given domain. This process relies on cognitive interviews with subject matter experts as to the goals of the task, which can be further divided into sub-goals. For example, a major goal in

---

[4] https://www.fusionproc.com/

driving could be to get to the destination safely and a sub-goal could be to avoid road hazards (Bolstad, Cuevas, Wang-Costello, Endsley, & Angell, 2010). However, driving operations in different industries, such as autonomous barges or remotely controlling a forklift in a warehouse, may diverge so completely that a different GDTA for driving would be required in each. This suggests that SA needs during remote operation will differ to SA requirements during in-situ driving, meaning that standard measurements to assess SA in driving may not be appropriate.

1.2 Implementation of SAGAT in other research studies

SAGAT is an influential tool but has not always been appropriately applied in the research world to investigate how drivers develop SA. Beyond the use of a freeze and probe approach, there is no agreed SAGAT methodology, nor a standardised method of data collection. Studies that have adopted SAGAT to measure driving SA have typically measured people's awareness in a range of different ways such as using a different number of queries to measure SA (Ma & Kaber, 2007; Scholtz, Antonishek, & Young, 2005), using a reconstruction task or a recognition task (Franz, Haccius, Stelzig-krombholz, Pfromm, & Kauer, 2015; Gugerty, 1997) or using a verbal recording in response to SA prompts whilst driving a Tesla in real time (Endsley, 2017). A meta-analysis of 243 papers which used SAGAT found that some researchers had developed queries that were not relevant to the SA requirements for the operational role or had failed to conduct a GDTA to determine their questions (Endsley, 2020). This limits the conclusions that we can draw about how SA is acquired and maintained in driving situations.

Within the design of SAGAT, there are concerning issues with the structure and design of the queries. Firstly, the three levels of SA queries are not measured equally; there are substantially fewer queries measuring comprehension and projection than perception suggesting results may be skewed in favour of Level 1 performance. Designing perception questions is a simpler task, as defining comprehension or projection in action is more challenging. Yet Level 3 projection has been cited as the highest level of understanding and skilled operators rely most on Level 3 SA to anticipate future events (Endsley, 2017). This lack of emphasis on equal assessment of SA at all levels is a concerning oversight of the tool; for an RO, perception of the remote environment will be vital, but they will need to understand how it feeds into what will happen next.

Secondly, some queries combine perception, comprehension and projection as "a parsimonious way to collect SA data", for example asking about objects that represent a hazard to show Level 1 SA knowledge of the *existence* of the hazard and Level 3 SA of the *risk* the hazard presents. Yet the 'blurring' of the levels within the questions such as this means that we cannot know which level of SA is missing if the answer is incorrect (Endsley, 2021, p. 23). To determine whether ROs have sufficient SA training to drive a vehicle remotely, we will first need to classify what remote SA knowledge is required to carry out the task but then need measures which can accurately test if that level of awareness has been achieved.

There is disagreement over the interpretation of what information constitutes the different levels of SA which, in turn, influences how responses to queries are recorded. For example, it is unclear whether the analysis of risk might belong in Level 2 Comprehension (understanding the risk presented by say, a pedestrian on the side of the road) or in Level 3 Projection (predicting the level of risk that that particular pedestrian poses, for example including a consideration of the fact that, if it is a child, they may be more likely to run into the road). Similar discrepancies exist for queries which ask participants to calculate distances; Bolstad et al. (2010, p. 5) judges "Distance to next turn on route" as Level 2 SA but Kaber, Zhang, Jin, Mosaly, & Garner (2012, p. 60) ask "When will we reach to the next turn indicated on the map?'' as a Level 3 SA query. It seems important that the separable components of an overall performance measure are kept mutually exclusive and unambiguous, in order to ensure that certain errors are not double counted due to their presence in more than one component.

Additionally, the language used in the queries is often open to interpretation. For example, Scholtz et al. (2005, p. 453) used a computerised multiple-choice interface to ask participants to complete a statement assessing Level 1 SA, *"The vehicle situation is…"* with the options, "*Normal/Cautionary/Dangerous/Don't know*". It is not known whether all participants would consistently apply the same interpretation of "cautionary" versus "dangerous".

Finally, research studies have additionally varied in their approach to scoring SAGAT responses, complicating comparisons between findings across studies. Endsley (2021) has recently provided clearer instructions of how to administer SAGAT with the aim of reducing errors and misinterpretation of the administration and design of SAGAT. She advises that a set of (say 20) queries, based on the GDTA, should be designed that variously measure perception, comprehension or projection and that can be objectively scored as correct or incorrect. At each freeze of the simulation, half of the queries should be asked using random sampling – they may or may not apply to the halted scenario. Each query should be asked/sampled at least 30-60 times in each experimental condition to have sufficient statistical power. Each query should then be scored separately (% correct), and any missing answer should be scored as incorrect. Some methodologies combine the queries to give a total SA score (Ma & Kaber, 2007) or have calculated three combined scores that represent Levels 1, 2 and 3 SA (Endsley, 1988a; Franz et al., 2015; Lo, Sehic, Brookhuis, & Meijer, 2016). Endsley has concluded that combining all the queries as an aggregated SA score or scoring each level of SA separately results in less sensitivity than reporting SA for queries individually. Taken together, these weaknesses point to a need for new approaches to measuring driving SA.

1.3 Aims of the study

The aim of this study is to develop a new approach to measuring driving SA that is appropriate for RO contexts. RO training should ideally draw on real-life examples of the types of use case in which they will be asked to provide assistance to AVs; only then can we evaluate whether they are sampling relevant SA information from the remote scene. In previous research (see Mutzenich et al, 2021), we investigated how participants build up mental representations of naturalistic remote driving scenes, simulating remote SA by using real footage filmed from a driving perspective, and unexpectedly asking participants to respond to questions at the end of the video. This uncovered, in its most basic form, what people ‘see’ in a remote scene when they are not constrained by rigid questioning, enabling the construction of a Taxonomy of Situation Awareness (TSA) that is directly applicable to remote driving contexts (see Appendix 1). This paper aims to address the lack of standardisation in previous quantitative driving-related SA measures by providing a new complete, open-source set of 16 videos of remote driving scenes[5] with accompanying queries for all SA levels.

This study followed a three-part process; first, a pilot study was conducted to trial the new remote SA measure and revise the queries on the driving videos, reported here in Section 3.1. Secondly, an experiment was carried out (*n= 94*) utilising the SA measure to investigate whether the presence of additional sources of visual and audio information improve SA. The results of this experiment are reported in Paper 4. Thirdly, an exploratory factor analysis was conducted to validate the SA measure using the datasets from the experiment, reported here in Section 3.6.

2.0 Development of materials

2.1 Creation of driving videos

To create the driving videos, researchers drove a 2016 petrol Dacai Duster 4x4 (manual right-hand drive) with two cameras fixed to the car recording continuously. Film footage was recorded on three separate occasions to develop materials that covered different times of day, traffic conditions, weather conditions and potential hazards. Video and audio recordings of the forward and rear views were filmed using two GoPro6 cameras mounted using the GoPro Suction Cup Mount. Two lenses captured the video with a spherical field of view (FOV) of 360 degrees (Sensor sizes are 1/2.3", with aperture of f/2.8). The front camera was mounted 60cm from the end of the bonnet and 50cm lateral from the widest point to ensure that it was not obstructing the driver's view and that the end

---

[5] https://github.com/Anthrometric/remote-operator.git

of the bonnet was not in shot. The rear camera was mounted 20cm up from the bottom of the rear-view window and 50cm lateral position, pointing at the road behind the vehicle. Both camera angles were verified from the driver's position inside the car to confirm that they accurately reflected the available forward-facing and rear-view perspective.

Recording started as soon as the driver released the handbrake and was controlled using a mobile phone link to the cameras. Videos were recorded at 60fps with a resolution of 1920 x 1080pixels. Sixteen separate videos were created by editing the total footage using Camtesia©.[6] They were converted from 60fps to 24 fps using Handbrake.fr for web optimisation. Eight of the videos in this study were shortened versions of the videos used in Mutzenich et al (2021). These videos demonstrated a range of road types with varying speed limits between 30-70 mph such as motorways, A and B roads, rural roads and residential areas. The videos ranged in size from 3.12 Mb (Fog video) to 19.15Mb (Pedestrian video) presented on an 811 x 456.188 screen placement holder.

Audio and video streams were separated, and the volume levelling of each clip was reduced to 88% of the original sound level. Camtasia's visual and audio effects library was applied to fade into the picture and audio at the start and end of each video.

2.2 Creation of rear-view and audio manipulations

Four versions of each video were created (shown in in Table 1) by either including or excluding the rear-view footage and the audio feed, giving 64 possible combinations of the sixteen videos. This experimental manipulation was used to test the experiment carried out in Paper 4 (n= 94) investigating whether the presence of additional sources of visual and audio information improve SA. The quantitative results of this experiment are reported in Paper 4.

Table 1 Description of the experimental conditions with abbreviations

Condition label	Description
No audio, no rear-view (NANR)	Forward facing video only with *no audio*
Audio, no rear-view (ANR)	Forward facing video only *with* audio
Rear-view, no audio (RNA)	Rear and forward-facing views with *no audio*
Rear-view, audio (RA)	Rear and forward-facing views *with* audio

---

[6] (www.techsmith.com, Version 2018. 0. 6).

The rear-view footage was rotated 180^0 to give a mirror image so that participants were given the impression of a rear-view mirror reflecting the road behind. This ensured that, in the video presentation, cars passing the ‘driver’ appeared on the right side of the road. The rear-view video was presented in a similar location as one would find a rear-view mirror in a vehicle (to the left of the driver in a right-hand drive car). It was set at a ratio of 30.2 % of the total forward facing video, positioned in the top left corner of the screen. This is based on the approximate proportional size of a rear-view mirror to a car windshield (see Figure 1 for examples of each video type). All videos were checked to ensure that the rear video feed did not occlude information in any of the sixteen separate videos, such as road signs.

Figure 1 Visual differences between rear-view absent (left) and rear-view present (right) conditions.

2.3 Development of SA questions

SA question design was based on previous research which took a novel approach to the SA Global Assessment Technique (SAGAT), employing a qualitative verbal elicitation task to investigate what people report from a remote scene when they are not constrained by rigid questioning (Mutzenich et al., 2021). This enabled the construction of a taxonomy of SA (TSA) in remote driving contexts (see Appendix 1). The TSA was used to design questions relevant to each video which could be answered in each condition (i.e. there were no questions related to information only visible from the rear view feed which would be unanswerable in the rear-view absent conditions nor questions relating to the audio feed which would be unanswerable in the audio absent conditions). Each question was designed to measure a particular level of SA (perception, comprehension, and prediction) and there were an equal number of questions presented for each SA level. Table 2 demonstrates the range of question types, measurement and examples of each type of question.

Table 2 Type of SA question with examples

SA question type	What information the question aims to measure	Categories for question design from TSA	Example question
Perception	Visual information visible in the remote scene	Traffic signs Colours of traffic lights Type of road or area road markings	"What colour were the traffic lights?".
Comprehension	Understanding of the remote scene	Road/highway Driver's perspective Other drivers/vehicles Pedestrians/other road users Current hazards	"Why are you waiting behind the blue car?"
Prediction	Ability to use information in the remote scene to make future predictions	Driver future action/s Other vehicle/road user predicted actions Changing road environment Predicted hazards	"What direction will you go in the next few seconds?"

Perception questions related to visual information visible in the remote scene such as traffic signs, colours of traffic lights, type of road or area and road markings for example "What colour were the traffic lights?". Comprehension questions measured participants' understanding of the scene, either from the driver's perspective, or from that of other road users. Endsley's SA model uses the term 'projection' to refer to Level 3 SA, yet this approach is commonly based on factual calculations not within the domain of the average driver (remember that Ensley's original research used military pilots adept at interpreting data from multiple instruments). Instead, we use the term 'prediction' to refer to Level 3 SA within the context of remote driving SA, as predictions are based on subjective analysis of known external factors, which may change and are uncontrollable but are likely to happen in the driver's experience. Prediction questions required participants to suggest what may happen in the next few seconds based on the ending of the video, for example how the road layout ahead may change.

3.0. Method

3.1 Pilot study

Twelve participants (9 females, 3 males) between 25 and 49 years of age (M^{age} = 34.9, SD^{age} = 8.77) with an average of 15 years driving experience (SD = 10.0) took part in the pilot study. Participants were equally likely to report driving every day (4), a few times a week (4) or once a month (4). We used Gorilla (www.gorilla.sc), an online cloud software platform, to host the study which took thirty minutes to complete using a desktop computer. Participants were informed how to withdraw from the study (by closing their browser window) and gave full informed consent before taking part. All procedures were reviewed and approved by the Royal Holloway Research Ethics Committee.

3.2 Design

The pilot study employed a 2x2 within-subjects factorial design: Audio (audio present, audio absent) x Rear (rear present, rear absent). We presented 16 videos counterbalanced across four conditions. The presentation of each video was randomly allocated to one of the four conditions.

3.3 Measures

Each video had three questions individually designed based on the events of the video, each of which was intended to get at one of the three SA levels. Correct answers were agreed by the research team from viewing the videos and participant responses were blind coded by the lead researcher. The maximum score in each condition was 12, giving a total possible score of 48 summed overall.

3.4 Procedure

Participants watched 16 videos of driving scenes with 4 videos in each condition and were asked three separate questions at the end of each video designed to measure each level of SA (perception, comprehension, and prediction) entering their answer in a text box. The order of the three questions was randomised to prevent practice effects. Figure 2 shows the complete procedure for the study.

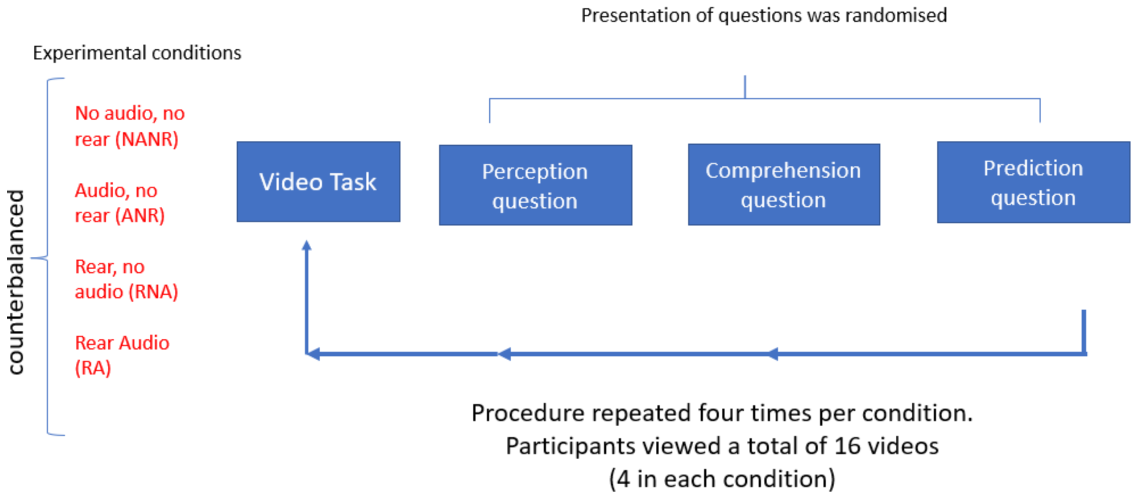


Figure 2 Pilot study procedure

4.0 Results

4.1 Modifications to SA questions

The pilot study allowed us to check whether the questions worked as expected or whether there were errors in question design or structure which could affect the validity of the SA findings. This allowed us to rephrase or change questions to limit miscommunication where necessary. The next section outlines the modification process in more detail for each of the SA question types and provides a record of all question changes (see Table 6 for all modifications).

4.1.2 Floor and ceiling effects of SA questions

The research team reviewed all participant responses to the pilot SA questions and determined what would be accepted as a correct answer, establishing a tolerance margin to participants' written responses for each question. For example, for the question, *"What were the pedestrians on the right-hand side doing with their car?"*, vague responses such as 'cleaning the window' required post-hoc judgements as to whether they came within tolerance of the correct answer of de-icing the window (this example did not).

It was apparent that many of the SA questions had design weaknesses. Floor and ceiling effects were considered present if less than 12.5% of participants achieved the lowest possible score overall or more than 83% achieved the highest possible score overall for that SA Level. Table 3 shows the floor and ceiling effects for each type of SA question.

Table 3 Details of floor and ceiling effects for each type of SA question

Type of SA question	Floor effects	Ceiling effects
Perception	1	6
Comprehension	1	7
Prediction	1	3

Many of the SA perception questions were too easy as they related to unchanging information directly in the front view during the duration of the video. We altered the wording in some questions to refer to information that was only visible in the last few seconds of the scene, to ensure that the SA Perception was being measured at the time of the 'freeze' rather than in relation to information that may have been perceived and stored in memory prior to that.

The answers to some questions were too obvious. For example, *"Why might the sun present a hazard in the next few seconds?"* could be answered even without seeing the video! Whereas 'Why should you leave a large space between yourself and the car in front for the next few minutes?' was more opaque and required comprehension that the sun glare meant that it would be difficult to see the brake lights of the car in front, so these types of questions were reworded (see Table 4 for a complete record of all question changes and Appendix 1 for the final modified SA questions).

4.3 Ambiguity in designing SA questions

Analysing performance question by question illustrated where there were areas of ambiguity. There were some generalised edits that were needed to make the questions clearer. For example, some participants were unsure about the temporal nature of the SA question, so they included information from the start of the video rather than just the last few seconds. To address this, we added "*at the end of the video*" to most questions. Some of the road terminology used, although correct, was confusing, as many people did not know what a motorway gantry is or what 'trajectory' referred to in certain contexts, so the wording was changed in these cases. Some questions were reworded to make the meaning clearer (e.g., changing *"driving action"* to *"driving manoeuvre"* as this requires specifics of the actual driving task). Any questions that could be answered yes/no and thus had a 50% chance of being correct were reworded as a 'what' or 'how' question.

Some questions did not have definitive answers. For example, *" Are you likely to accelerate in the next few seconds?"* could be answered positively or negatively without either being objectively incorrect, seeing as the question referred to a future driver action that would not be predictable with absolute certainty. Questions of this type were amended as necessary. The fact that changes of this nature to the pilot questions were necessary highlight that there are many barriers to effective SA that come from language communication, even though all participants were native English speakers.

We also wanted to ensure that there was an equal representation of all the sub-sections of the TSA, so we added in perception and comprehension questions relating to road signage (TSA Comprehension 1.1) as this was judged not to be directly assessed in the original questions.

Some questions were deemed to be too generic, for example, '*How will the road ahead change in the next few seconds?*'. In its place, we adapted the question to be more related to the specific context of the video i.e. *'Why should you change down to a lower gear in the next few seconds?'*, which was related to the road layout visible in last ten seconds of the video, as the vehicle was about to go up a hill. This question now demonstrated understanding of defensive driving which requires drivers to anticipate the driving actions that they may need to take in the next few seconds related to the possibilities of how the current scene may unfold and relates directly to SA Prediction skills.

Some questions were altered as it was decided that they had two answer options which were both true. For example, in one video *"What were the pedestrians on the right-hand side doing with their car?"* could be answered with reference to either of two pedestrians – one of whom was de-icing the window (originally determined to be the 'correct' answer) but the other of whom was getting into the car. Although one answer was deemed 'correct' the other answer could be considered a reasonable response, so the wording was refined such that only one correct answer was possible: *"What was the driver standing on the road on the right-hand side doing to their car?".*

4.4 Blurred boundaries between SA levels

Some questions were changed from comprehension to perception questions, reflecting the fact that the boundaries between the different 'levels' of SA are often difficult to define. In Mutzenich et al. (2021), we discussed the fact that the TSA revealed blurring of the traditional boundaries assumed to exist between comprehension and prediction and it is possible that this is true of perception and comprehension too. We asked an independent group of people to evaluate whether they felt the SA questions reflected perception or comprehension and adjusted accordingly. For example, one question *"What was unusual about the bus stop you passed on your left?"* (answer – a car was parked in it) required understanding of the fact that it is inappropriate for a car to be parked in a bus stop but also required visual perception and attention to notice it was there. This question was judged to be measuring comprehension as it required knowledge that cars are not supposed to park in bus stops rather than just perceptual awareness of the car. Yet some people answered literally that it was a wooden hut demonstrating SA Perception. This was true, and it could also be seen as unusual as in the UK bus stops are more frequently metal constructions or a single bus stop sign, so this question was eventually replaced. This highlights the challenge of targeting SA questions at particular SA levels and the importance of not assuming that participants will necessarily adopt the same interpretation as experimenters. A remote scene provides an extensive choice of potential target answers, so deciding on SA questions that only have one correct answer is challenging, yet a necessity when scoring SA performance (Mutzenich et al., 2021).

Table 4 Complete modifications to SA questions from results of the pilot study

Video scenario	Perception question	Perception Revised Question
A road #2	What type of vehicle were you travelling behind when the video stopped?	What on-road sign in your lane had you just passed at the end of the video?
Sun	What type of road are you travelling on?	What colour are the lines on the side of the road that you are travelling on?
Urban	In what direction did the red car in front of you turn off?	What type of vehicle was waiting to pull out at the traffic lights you passed on the left?
Pedestrian	How many cyclists were in the scene?	What natural feature is to your right on the street you are driving on?
Fog	Which side of the road was the traffic sign for a village located?	At the end of the video, what type of business did you pass on your right-hand side?
Roundabout	What season was it in the video?	What was unusual about the bus stop you passed on your left?
Winter	In which direction did you turn out of the junction?	What was the driver standing on the road on the right-hand side doing to their car?
	Comprehension question	**Comprehension Revised Question**

A road #1	Why were you stopped behind the grey van?	Why was traffic still passing on the right while you were stationary?
Motorway #2	Which lane are you currently driving in?	At the end of the video, what is the relative position of the nearest car to you?
Residential #1	Why are you waiting behind the blue car?	Why did you need to leave a gap between the parked blue car and yourself at the start of the video?
Residential #2	Why did the van slow down?	Why was it necessary to leave extra space whilst overtaking the van?
Rural #1	What action should you be taking with regards to your speed?	At the end of the video, what was the on-road sign instructing you to do?
Sun	What do the double yellow lines on the side of the road mean?	Why would it be a problem if you needed to stop to make a delivery on this road?
Urban	What are the locations of the cars surrounding you?	At the end of the video, how many lanes could you use to go straight ahead at the traffic lights?
Pedestrian	What was the location of the second cyclist when they stopped?	Why did the second cyclist (on your right) stop?
Fog	What do the dotted white lines on the road indicate?	According to the last sign you passed, how will the road change ahead?
Roundabout	What was unusual about the bus stop on your left?	What check did you need to make before you entered the roundabout without stopping?
Winter	What were the pedestrians on the right-hand side doing with their car?	What are the weather conditions in this video that should influence your speed?

	Prediction question	**Prediction Revised Question**
Motorway #1	How long before there is a stopping bay ahead?	Why will it be difficult to pull over on this section of road if you develop a problem with your car?
Motorway #2	What is the position/location of the nearest car to you?	At the end of the video, what manoeuvre may the nearest car to you make next?
Residential #2	How will the road ahead change in the next few seconds?	Why should you change down to a lower gear in the next few seconds?
Rural #2	What will be your future trajectory on this road?	What potential hazard should you be aware of on the road ahead, as a result of the type of work being carried out by the vehicle you just passed?
Sun glare	Why might the sun present a hazard in the next few seconds?	Why should you leave a large space between yourself and the car in front for the next few minutes?
Pedestrian/cyclist	What direction will you go in the next few seconds?	How may the absence of central white lines on the road present a hazard in the next few seconds?
Fog	Are you likely to accelerate in the next few seconds?	Why might it be dangerous in the current weather conditions if the car ahead brakes suddenly?
Roundabout	Will you have to stop suddenly if the car in front of you brakes?	Why might it be likely that you encounter cyclists on this road up ahead?
High traffic density	Will you have to stop again in the next few seconds?	At the end of the video, what is causing the traffic to still be stationary?
Junction	As you turned out of the junction, what was ahead that could present a hazard?	As you turned out of the junction, what was immediately in front of you that could present a hazard?
Winter	What may cause you to brake in the next few seconds to avoid a collision?	What may cause you to brake in the next few seconds to avoid a collision?

5.0 Validation of the SA measure

To identify the underlying constructs of the instrument, we conducted an exploratory factor analysis. An empirical study (n=96) is reported in Paper 4 which applied the SA measure described here, using the same experimental procedure and driving videos employed in this study, to investigate whether the presence of additional sources of visual and audio information improve SA. Datasets taken from Paper 4 results are utilised in the next section to validate whether the questions successfully measured remote SA.

5.1 Justification of analysis

Exploratory factor analysis (EFA) identifies underlying constructs in measurements scales, which in the current study relates to whether the questions accurately measure SA (DeCoster, 1998). Common relationships between categories of questions should mean that they load onto the same factors, making EFA an appropriate tool for measuring the validity of the SA measurement scale (Williams, Onsman, & Brown, 2012). Traditional EFA measures require ordinal and continuous data with corresponding assumptions that the dataset follows a normal distribution, for example Likert scale responses to measure the strength of participant attitudes. However, the answers to the questions measuring SA Perception, SA Comprehension and SA Prediction for the driving videos in this study were binary coded, as either correct or incorrect, making the dataset unsuitable in its original format for factor analysis. In cases such as these, tetrachoric correlation techniques can be used to convert the dataset into a binary matrix where the correlations for each pair of variables in a data frame are estimated as if they are continuous data.[7] Thus, dichotomous datasets are treated as *underlying* continuous variables with assumed normal distributions which permits factor analysis on categorical variables, and this is the approach used in this study (Barendse et al., 2015).

There is considerable disagreement in the literature as to the minimum sample size recommended for EFA, ranging from at least 500 subjects to as few as 55, depending on the proportion of variance explained by latent factors or, alternatively, variable-to-factor ratios (Pearson & Mundfrom, 2010). However, this does not necessarily restrict EFA being used to explore patterns underlying a dataset to develop new theories, provided results are treated with caution and tested on larger samples in future iterations of instrument design (Knekta, Runyon, & Eddy, 2019). As a result, the relatively small dataset (n= 94) in this study, compared to what is usually recommended for this type of analysis, is considered sufficient as the study intends to develop a new approach to measuring driving SA which requires an understanding of the underlying constructs of the measure. To safeguard the legitimacy of our approach when fitting the model, we followed recommendations that each factor should be loaded by at least three variables, each variable should load on only one factor, factors should have internal consistency reliability ≥.70, and there should be a theoretical justification for factor labels (Watkins, 2018).

---

[7] https://www.statisticshowto.com/tetrachoric-correlation/

5.2 Exploratory factor analysis

An exploratory factor analysis was conducted on 48 items with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin measure (KMO = .6) endorsed the sampling adequacy as 'mediocre' but, as KMO values for individual items were all above the required limit (.5), it was judged as suitable for EFA (Williams et al., 2012). Correlations were all below r = +/- .90 indicating that there was no multicollinearity. We used the psych package in R to conduct maximum-likelihood factor analysis on a polychoric covariance matrix. Parallel analysis suggested that eight factors should be retained but low factor loadings in this model were below the KMO boundary. Examination of the scree plot suggested the extraction of either two or four factors with eigenvalues over Kaiser's criterion of 1. We retained four factors as this model explained a higher variance (36%) than a two-factor model (24%). Appendix 1 shows the factor loadings (>.5) after rotation, which suggests that the SA measure has four underlying constructs. Factor 1 represents spatial and environmental awareness, factor 2 represents anticipatory hazards, factor 3 dynamic driving actions and factor 4 represents other road users. Internal consistency of the criterion set was evaluated by Cronbach's alpha coefficient at α= 0.832, 95% CI[0.734, 0.896].

5.3 Justification of factor labelling

Factor 1 was composed of SA queries which required spatial awareness within the remote scene. For example, the highest loading item (.73) asked "*What may cause you to brake in the next few seconds to avoid a collision*?" which requires participants to notice what is currently going on around the ego vehicle and a calculation of what is too close. Further items required three-dimensional awareness of the remote scene, for example two items asked about signs they had already passed in the video giving information about the road layout ahead, other items required locating perceptual information either to the right or left of the remote vehicle or related to their on-road positioning. This factor was classed as 'spatial and environmental awareness'.

Factor 2 related to an SA need to analyse the risk in the driving videos, either through direct questions which asked about traffic warning signs and the hazards posed by a particular stimulus in the remote scene or more oblique references to 'risky' behaviours. For example, the item " What was the driver standing on the road on the right-hand side doing to their car?" contains information that someone is standing in the middle of the road which would be a potential hazard, even if the point of the question is asking about their action towards their car. One item, "When you turned out of the junction, what was immediately in front of you that could present a hazard?", loaded highly onto Factor 2, but also onto Factor 1, as the first part of the question referred to what was in front of them, but the second part of the question related to awareness of the hazard it presented. A few items referred to emerging hazards such as "*What are the weather conditions in this video that* *should influence your speed*?" which may be examples of defensive driving, drawing on anticipatory responses to current information rather than relying on what is currently presented in the remote scene (Walker, Stanton, Kazi, Salmon, & Jenkins, 2009). We termed this factor ‘anticipatory hazards’ to incorporate both present and future hazard awareness.

Factor 3 consisted of items which referred to driving manoeuvres carried out by both the participant in their role as the driver of the ego vehicle, or by other drivers in the remote scene. These manoeuvres included pulling out of junctions, stopping and starting and changing gears. The item that asked, "*Why did you need to leave a gap between the parked blue car and yourself at the start of the video*?" related to a video which showed the ego vehicle waiting on a residential road with parked cars on one side, so the question anticipated awareness of the manoeuvres that oncoming traffic would need to carry out to slip into the space between the final parked car and the remote ego vehicle. We labelled this factor ‘dynamic driving actions’.

Finally, although fewer items loaded onto Factor 4, the contribution of each variable to the factor was very strong. The items "*What was the purpose of the vehicle parked on the grass verge*?" and "*What potential hazard should you be aware of on the road ahead, as a result of the type of work being carried out by the vehicle you just passed*?" loaded above .7, meaning that they are well explained by the factor. Both these items related to the same video, which showed an abandoned council work vehicle used for clearing highway debris parked on the side of the road which implied that the driver was potentially up ahead as a pedestrian. The third item related to the, also unseen, presence of cyclists on the road ahead suggesting that this item was related to an awareness of non-motorised vehicles and pedestrians also sharing the highway. We termed this factor ‘other road users’.

5.4 Comparison of factor labels with the categories of the Taxonomy of Situation Awareness in Driving

Loadings on all factors ranged across three types of SA question, which is illustrative of the blurred boundaries between the three SA levels. For example, spatial and environmental awareness requires a complex interaction between understanding perceptual information collected from the remote scene and analysing its significance in current and future states of driving. This echoes the qualitative finding from Mutzenich, Durant, Helman, & Dalton's (2021) previous study (reported in Paper 2) that SA in driving contexts is likely to involve a process of combining the SA levels in parallel rather than serially.

We would expect to see evidence from the factor analysis that the underlying constructs of the SA measure that we have developed to assess remote driving SA corresponds with themes from the taxonomy that we used to design the SA questions to accompany the videos. For example, the themes 'Road/highway' and 'Driver's perspective' in the SA Comprehension category and 'Changing road environment' in the SA Prediction category within the TSA, contain information relating to the road layout and the driver's relative position to other cars which match the spatial awareness labelled as Factor 1. This agreement is also true for other factors with corresponding categories within the TSA, shown in Table 5. In sum, factor structure seems to correspond fairly well with previous categorisation of the SA concepts that underpin remote driving SA.

Table 5 Comparison of TSA categories and factor labels for exploratory factor analysis (n=94) of a novel SA measure

TSA Category and theme	**Factor label**	**SA Question**
SA Comprehension 1.1 Road/highway 1.2 Driver's Perspective **SA Prediction** 2.1 'Changing road environment'	**Factor 1** **Spatial and environmental awareness**	*What may cause you to brake in the next few seconds to avoid a collision?* *There is no junction coming up, so why did you move over to the left hand lane?* *According to the last sign you passed, how will the road change ahead?* *What did the traffic sign you passed on your left indicate was half a mile ahead?* *What type of area are you currently waiting in?* *At the end of the video, what road feature is signposted as coming up on the road ahead of you?* *What colour were the traffic lights in your lane when the video stopped?* *Why was it necessary to leave extra space whilst overtaking the van?* *As you turned out of the junction, what was immediately in front of you that could present a hazard?* *What natural feature is to your right on the street you are driving on?* *What was unusual about the bus stop you passed on your left?* *What colour are the lines on the side of the road that you are travelling on?*
SA Comprehension 1.6 Anticipatory hazards	**Factor 2** **Anticipatory hazards**	*What was the traffic warning sign on the motorway gantry you passed under?* *What was the driver standing on the road on the right-hand side doing to their car?* *What are the weather conditions in this video that should influence your speed?* *In this video, why did you wait at the junction before pulling out?* *Why might it be dangerous in the current weather conditions if the car ahead brakes suddenly?* *Why will it be difficult to pull over on this section of road if you develop a problem with your car?* *What check did you need to make before you entered the roundabout without stopping?* *At the end of the video, what is causing the traffic to still be stationary?* *Why did the second cyclist (on your right) stop?* *Why did the car turning into the car park not enter it immediately?*
SA Comprehension	**Factor 3**	*What driving manoeuvre will you carry out in the next few seconds?*

1.3 Dynamic driving actions of driver and other drivers/vehicles **SA Prediction** 2.1 Driver future action/s	**Dynamic driving actions**	*What type of vehicle was waiting to pull out at the traffic lights you passed on the left?* *At the end of the video, what manoeuvre may the nearest car to you make next?* *Why would it be a problem if you needed to stop to make a delivery on this road?* *Why did you need to leave a gap between the parked blue car and yourself at the start of the video?* *Why should you change down to a lower gear in the next few seconds?* *Why was traffic still passing on the right while you were stationary?*
SA Comprehension 1.5 Other road users/pedestrians **SA Prediction** 2.2. Other vehicle/road user predicted actions	**Factor 4** **Other road users**	*What was the purpose of the vehicle parked on the grass verge?* *What potential hazard should you be aware of on the road ahead, as a result of the type of work being carried out by the vehicle you just passed?* *Why might it be likely that you encounter cyclists on this road up ahead?*

6.0 Discussion

6.1 Comparison to previous literature

In the current study, we used the TSA to design and validate driving SA questions for 16 original driving videos. An exploratory factor analysis on the 48 items from the SA measure found four underlying constructs which we termed ‘spatial and environmental awareness’, ‘anticipatory hazards’, ‘dynamic driving actions’ and ‘other road users’.

Our findings broadly support Endsley's concept of SA (1988) as incorporating what is known about the environment, what is happening in it and what might change in the near future to build awareness. However, the results of our factor analysis suggest that focusing too narrowly on the divisions between the three levels of SA may lack construct validity in the context of driving situation awareness. Factor loadings demonstrated that the separate SA questions asked in this study on perception, comprehension and prediction were combined when building awareness of a remote driving scene, which suggests that other driving SA measurement tools intentionally designed to hold SA levels separate may be inappropriate.

Earlier research using SAGAT to measure SA in driving (such as Bolstad (2000); Kaber et al (2012), Kaber & Ma (2005); Scholtz (2005); Franz, (2013); Gugerty, (1997); Endsley, (2017)) demonstrated inconsistencies in the measurement, design and structure of SA queries, in part due to the challenging nature of operationalising perception, comprehension and projection into measurable constructs. Our research also exposed the challenges of designing questions which only measure one level of SA even after the extensive redrafting and refinement process conducted in this study, as was shown by the EFA revealing a factor structure that did not appear to map directly on to the SA levels. Yet, to our knowledge, no studies which employed SAGAT reported amending questions based on pilot study feedback nor have they reported addressing floor or ceiling effects in responses to specific queries as we did in this study. We also allowed participants to write their answers in full text, minimising the structural and linguistic deficits of multiple choice used by previous SA studies into driving such as Scholtz, Atonishek & Young (2005). In summary, the current work has attempted to go beyond many of the existing approaches for measuring SA, by using iterative qualitative and quantitative assessment to determine whether the questions are operating as intended on a number of levels.

Previous findings from Mutzenich, Durant, Helman, & Dalton (2021) (reported in Paper 2) that SA levels are likely to be achieved in parallel rather than serially were in line with the findings of this study. During the development of the TSA, driving operations dominated the qualitative descriptions of what was happening in the driving videos, together with the presence of pedestrians and other road users, and this is echoed in the results in this study. In contrast to research (Endsley, 2020) which discovered some researchers develop SAGAT queries that are not relevant to the SA requirements for the operational role, we presented the videos from the perspective of the visual feed of a remote operator to get a sense of the issues that will be involved when attempting to build awareness second-hand from a remote scene and we designed the SA questions in this study specifically to match the type of information reported by participants in Study 1. We used real footage filmed during live driving, relinquishing some design control over what happened in the videos but giving a more realistic experience. We argue that our stimuli typify the range of unexpected situations which an RO might have to contend with on a moment-to-moment basis, on different roads, weathers and potentially hazardous developing situations. Although we are aware of the limitation of applying the results from this more passive task of watching pre-recorded videos to the more active task of carrying out live remote operations, this method is also more realistic than simulations and mirrors some of the sensory restrictions which would characterise an RO’s experience in building and maintaining SA. Consequently, we believe that the measure is relatable to the SA requirements of an RO.

6.2 Limitations of the study

We learned from the pilot study that participants’ familiarity with the rules of the road/highway code cannot be assumed and in fact seem to be debatable in some cases! Although an RO would have to be trained to a standard that complies with all highway regulations, for the purposes of this study with the general population, we removed questions which required precise knowledge of highway regulations. This is illustrative of the differences between regular drivers as opposed to operators in more regulated occupations, which again emphasises the importance of acknowledging the relative lack of expertise of most regular drivers in comparison with the military pilots with whom the SAGAT was originally developed. For example, research has shown that novice drivers can take eight seconds longer to gain SA than more experienced/older drivers (Wright et al., 2016) although this was measured in Level 3 autonomous driving tasks (where the in-situ driver is expected to take over from the automated system) not during remote operations. Nevertheless, in the future it is likely that RO roles will be given to experienced drivers, so while in this study we required a minimum of three years of driving experience, in future studies it may be advisable to use professional drivers as the sample to mirror the probable characteristics of this labour force.

The pilot study provided a strong basis for refining the phrasing of the SA questions. This work was also useful in assessing the suitability of the online approach to data collection. Using the analytics from the online platform, we could see that all the videos played without errors or lagging, and participants' feedback in the debrief also confirmed that the size and visibility of the videos were appropriate to enable all the questions to be answered. The online methodology used in this study to assess driver SA shows that using online training to identify and encourage the SA skills necessary for carrying out RO tasks could be a viable solution to coach remote operations workers who may be employed in multiple global locations.

One of the limitations of the EFA technique used in this study to validate the SA measure is that naming the factors can be subjective, for example factor names may not accurately reflect the variables within the factor, particularly in the case of when variables load onto more than one factor. We used the TSA to corroborate the factor labels to provide a theoretical justification for the latent constructs identified by the factor analysis, but this could be open to researcher bias as we also authored the earlier paper. The small sample size (n=94) is also marginally lower than is recommended by some researchers, but this was determined by a power analysis for the study parameters reported in Paper 4 and provided an opportunity to validate the measure using the datasets collected in live testing. Furthermore, we made judgements about how many factors to retain, between 2, 4 and 8, to strike a balance between underestimation and over determination, but again this could be open to interpretation (Knekta et al., 2019). However, we satisfied the criteria from evidence-based guidance (Watkins,2018), that factor loadings should have a minimum of 3 variables, load onto only one factor, with an internal consistency reliability ≥.70, and show transparent theoretical justification for factor labels. As a result, although we advise caution in drawing fixed conclusions about the nature of remote SA from this measure, we consider it a strong step towards developing a more sensitive instrument to measure driving SA in remote environments.

6.3 Future research recommendations

Increasingly, transportation research is concentrating on questionnaire development and behavioural measures to understand driving, including the application of EFA techniques to validate their dimensionality (Ledesma et al., 2021). The purpose of conducting factor analysis is to understand how interactions operate between different items and constructs, and our research goes towards developing a validated tool to measure the SA needs of remotely operating a vehicle. However, the caveats raised above suggest that there is much work to be done in this area of research to increase the transparency and reliability of EFA practices and we remain open to new developments in this field, such as hybrid models between factor analysis and Bayesian analysis using researchers' knowledge as informative ''priors" (Ledesma et al., 2021).

There are several projects currently underway in the UK designed to understand every aspect of remote operation, including human factors considerations. One of the key contributions of this paper is therefore an open-source resource for measuring SA by using videos of remote driving situations. Our results suggest that we should identify the skills we wish ROs to develop and design training videos and accompanying questions to extract those aspects, rather than relying on all-purpose approaches across many domains, such as SAGAT. Recognising that remote driving SA is a multidimensional construct should motivate many novel approaches to SA research in the future as remote operation becomes a standardised and regulated certainty in our societies.

References

Barendse, M. T., Oort, F. J., Timmerman, M. E., Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 87–101. https://doi.org/10.1080/10705511.2014.934850

Bolstad, C., Cuevas, H., Wang-Costello, J., Endsley, M. R., & Angell, L. S. (2010). Measurement of situation awareness for automobile technologies of the future. *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*, *4970*(June), 195–213. https://doi.org/10.4271/R-402

DeCoster, J. (1988). An Overview of Factor Analysis. https://doi.org/10.4135/9781412984898.n1

Endsley, M. R. (1988a). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting*, *32*(2), 97–101. https://doi.org/10.1177/154193128803200221

Endsley, M. R. (1988b). Situation awareness global assessment technique (SAGAT). *Aerospace and Electronics Conference*, 789–795. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R. (2017). Toward a theory of situation awareness in dynamic systems. *Human Error in Aviation*, 217–249. https://doi.org/10.4324/9781315092898-13

Endsley, M. R. (2020). The Divergence of Objective and Subjective Situation Awareness: A Meta-Analysis. *Journal of Cognitive Engineering and Decision Making*, *14*(1), 34–53. https://doi.org/10.1177/1555343419874248

Endsley, M. R. (2017). *Autonomous Driving Systems : A Preliminary Naturalistic Study of the Tesla Model S*. https://doi.org/10.1177/1555343417695197

Endsley, M. (2021). *Situation Awareness Measurement : How to Measure Situation Awareness in Individuals and Teams*. Human Factors & Ergonomics Society.

Endsley, M, Bolte, B., & Jones, D. (2003). *Designing for Situation Awareness: An Approach to User-Centered Design* (2nd ed.; N. Y. T. & Francis, Ed.). https://doi.org/10.1201/9780203485088

Franz, B., Haccius, J. P., Stelzig-krombholz, D., Pfromm, M., & Kauer, M. (2015). Evaluation of the SAGAT method for highly automated driving. *Proceedings 19th Triennaial Congress of the IEA*, (August), 1–8.

Gugerty, L. J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. *Journal of Experimental Psychology: Applied*, *3*(1), 42–66. https://doi.org/10.1037/1076-898X.3.1.42

Kaber, D., Zhang, Y., Jin, S., Mosaly, P., & Garner, M. (2012). Effects of hazard exposure and roadway complexity on young and older driver situation awareness and performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(5), 600–611. https://doi.org/10.1016/j.trf.2012.06.002

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn’t fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE Life Sciences Education*, *18*(1), 1–17. https://doi.org/10.1187/cbe.18-04-0064

Ledesma, R. D., Ferrando, P. J., Trógolo, M. A., Poó, F. M., Tosi, J. D., & Castro, C. (2021). Exploratory factor analysis in transportation research: Current practices and recommendations. *Transportation Research Part F: Traffic Psychology and Behaviour*, *78*, 340–352. https://doi.org/10.1016/j.trf.2021.02.021

Lo, J. C., Sehic, E., Brookhuis, K. A., & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 325–338. https://doi.org/10.1016/j.trf.2016.09.006

Ma, R., & Kaber, D. B. (2007). Situation awareness and driving performance in a simulated navigation task. *Ergonomics*, *50*(8), 1351–1364. https://doi.org/10.1080/00140130701318913

Mutzenich, C., Durant, S., Helman, S., & Dalton, P. (2021). Situation Awareness in Remote Operators of Autonomous Vehicles: Developing a Taxonomy of Situation Awareness in Video-Relays of Driving Scenes. *Frontiers in Psychology*, *12*(November). https://doi.org/10.3389/fpsyg.2021.727500

Pearson, R. H., & Mundfrom, D. J. (2010). Recommended sample size for conducting exploratory factor analysis on dichotomous data. *Journal of Modern Applied Statistical Methods*, *9*(2), 359–368. https://doi.org/10.22237/jmasm/1288584240

Scholtz, J. C., Antonishek, B., & Young, J. D. (2005). Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *35*(4), 450–459. https://doi.org/10.1109/TSMCA.2005.850589

Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., & Jenkins, D. P. (2009). Does advanced driver training improve situational awareness? *Applied Ergonomics*, *40*(4), 678–687. https://doi.org/10.1016/j.apergo.2008.06.002

Watkins, M. W. (2018). Exploratory Factor Analysis : A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Williams, B., Onsman, A., & Brown, T. (2012). Exploratory factor analysis: A five-step guide for novices EDUCATION Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, *8*(3), 1–13. Retrieved from http://ro.ecu.edu.au/jephc/vol8/iss3/1

Wright, T. J., Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D. L., Swedish National, R., & Transport Research, I. (2016). *Are experienced drivers quicker to regain full situation awareness in scenarios involving transfer of control from the automation to the driver? 1*(3), 3p. Retrieved from http://vti.diva-portal.org/smash/get/diva2:924898/FULLTEXT01.pdf https://trid.trb.org/view/1454137

Appendix

Table 6 Factor loadings for each SA question

Factor		Spatial and enviromental awareness	Anticipatory hazards	Dynamic driving actions	Other road users
% of variance		.11	.10	.08	.06
SA Question type	*SA Question*				
Prediction	What may cause you to brake in the next few seconds to avoid a collision?	**.73**			-.13
Comprehension	There is no junction coming up, so why did you move over to the left hand lane?	**.65**	.28		-.22
Comprehension	According to the last sign you passed, how will the road change ahead?	**.63**	.13		.19
Perception	What did the traffic sign you passed on your left indicate was half a mile ahead?	**.62**	.13		
Perception	What type of area are you currently waiting in?	**.58**	-.11		.16
Prediction	At the end of the video, what road feature is signposted as coming up on the road ahead of you?	**.55**	.28	.33	.28
Perception	What colour were the traffic lights in your lane when the video stopped?	**.49**		.12	
Comprehension	Why was it necessary to leave extra space whilst overtaking the van?	**.45**	.36	.15	
Prediction	As you turned out of the junction, what was immediately in front of you that could present a hazard?	**.44**	**.59**	-.17	
Perception	What natural feature is to your right on the street you are driving on?	**.44**		.19	
Perception	What was unusual about the bus stop you passed on your left?	**.41**		.28	
Perception	What colour are the lines on the side of the road that you are travelling on?	**.40**		.11	.23
Perception	What was the traffic warning sign on the motorway gantry you passed under?		**.76**	-.62	-.19
Perception	What was the driver standing on the road on the right-hand side doing to their car?		**.71**		
Comprehension	What are the weather conditions in this video that should influence your speed?	.16	**.63**	.13	
Comprehension	In this video, why did you wait at the junction before pulling out?	.12	**.59**		-.19
Prediction	Why might it be dangerous in the current weather conditions if the car ahead brakes suddenly?	-.14	**.57**		
Prediction	Why will it be difficult to pull over on this section of road if you develop a problem with your car?	.29	**.49**	.14	.15
Comprehension	What check did you need to make before you entered the roundabout without stopping?	.36	**.44**	.39	.16
Prediction	At the end of the video, what is causing the traffic to still be stationary?		**.44**	.29	
Comprehension	Why did the second cyclist (on your right) stop?		**.42**		.29
Comprehension	Why did the car turning into the car park not enter it immediately?		**.42**	.35	.24
Prediction	What driving manoeuvre will you carry out in the next few seconds?			**.64**	-.11
Perception	What type of vehicle was waiting to pull out at the traffic lights you passed on the left?			**.62**	
Prediction	At the end of the video, what manoeuvre may the nearest car to you make next?	.20	.35	**.55**	
Comprehension	Why would it be a problem if you needed to stop to make a delivery on this road?	.22		**.48**	
Comprehension	Why did you need to leave a gap between the parked blue car and yourself at the start of the video?	.11		**.43**	
Prediction	Why should you change down to a lower gear in the next few seconds?		.18	**.43**	
Comprehension	Why was traffic still passing on the right while you were stationary?	.38		**.42**	
Comprehension	What was the purpose of the vehicle parked on the grass verge?	.39	.34	.12	**.76**
Prediction	What potential hazard should you be aware of on the road ahead, as a result of the type of work bei	.29	.33		**.70**
Prediction	Why might it be likely that you encounter cyclists on this road up ahead?	.21			**.45**
Comprehension	At the end of the video, how many lanes could you use to go straight ahead at the traffic lights?		-.11	.23	.37
Perception	At the end of the video, what type of business did you pass on your right hand side?		.33	.34	.35
Perception	What type of shop/s are you waiting by?	.17		.33	.23
Prediction	What are the two turning options ahead in your lane at the traffic lights at the end of the video?	.33	.21	.16	.12
Perception	What on-road sign in your lane had you just passed at the end of the video?	.18	.26	.19	-.17
Comprehension	At the end of the video, what is the relative position of the nearest car to you?				-.17
Perception	What type of traffic calming device was obstructing the road you were driving on?	.39	.33	.34	-.18
Prediction	How is the road layout ahead potentially hazardous?	.27	.28	.19	-.19
Comprehension	Why was the road you passed to the left blocked?	.22		-.12	-.21
Comprehension	At the end of the video, what was the on-road sign instructing you to do?	-.24		.15	-.37
Prediction	How may the absence of central white lines on the road present a hazard in the next few seconds?	.33		.28	-.42
Perception	What turning signal did the van in front of you give before it stopped?	.17	.15	-.30	-.62
Perception	Which exit did you take off the roundabout?	.24	.35	.36	
Prediction	Why should you leave a large space between yourself and the car in front for the next few minutes?	.34	.13	.32	
Prediction	What action might the van in front take which would require you to slow down in the next few secor	-.16	-.36	.22	
Perception	What position, relative to you, was the car waiting to pull out of the junction?	.27		.17	

nb. factor loadings of <.2 are supressed. Factor loadings >.40 are in bold

Paper 4

Influence on remote SA of two information sources presented in addition to the standard front camera feed

Abstract

There are several projects currently underway in the UK designed to understand every aspect of remote operation, including human factors considerations such as situation awareness (SA); what is known about the environment and what may change in it. This research investigated whether the provision of rear-view and/or auditory information could improve SA for remote driving scenes. Using the Taxonomy of Situation Awareness (TSA) developed in previous research (see Mutzenich et al, 2021), we designed SA questions for sixteen original driving videos. We presented the videos from the perspective of the visual feed of a remote operator to get a sense of the issues that will be involved when attempting to build awareness second-hand from a remote scene. We hypothesized that SA performance would be highest in the condition in which both the rear-view and audio feeds were presented, as this would provide the most information to assist in building awareness. We also expected participants' sense of presence to be highest in audio conditions due to the additional sensory information that was provided. We adopted a novel scoring protocol where only correct answers with high confidence in their response (selected 4 or 3 on a confidence scale) contributed to SA performance score. We consistently found worse SA performance in the presence (vs. absence) of the rear-view feed, which was unexpected as previous research had suggested that the additional rear-view information would be of benefit when building SA. Our research highlights that, depending on task goals for remotely operating automated vehicles, these information sources are not always as helpful as we might assume and may occasionally impair performance.

1. Introduction

A pervasive misconception in the transport industry is that automated vehicle transport can only be considered autonomous if there is no human involved at all. In reality, many problems can arise that would require a human operator to remotely assess and instrumentally correct or direct the automation, because autonomous vehicles (AVs) can sometimes fail to interpret even seemingly straightforward perceptual information correctly. These types of programming deficits are referred to as edge cases - random, unusual occurrences triggered when an AV cannot match the perceptual information with known datasets, due to confusing or partial stimuli (Mutzenich et al., 2021a). Consequently, a remote operator may occasionally be alerted by an AV to drop in and assess an edge case. The functional scope that a remote operator offers can span from remote assistance, such as interacting with passengers, through remote management where they merely advise the AV of how to proceed once they have reviewed the edge case, through to direct remote control, whereby they may have to take over dynamic driving control of the AV from a remote location (Mutzenich, Durant, Helman, & Dalton, 2021b).

To carry out any of these tasks, a remote operator will first need to acquire situation awareness (SA) of the remote scene. SA involves drawing on perceptual information (Level 1 SA "Perception") to make sense of what is happening in a remote scene (Level 2 SA "Comprehension") and what might change in the near future (Level 3 SA "Projection")(Endsley, 1988). A vital part of SA is "knowing what is going on around you", so the operator's physical absence will make the task more difficult as they will need to build up a mental representation of the remote environment primarily by monitor views of video feeds from the scene (Endsley, 2000: pg 3). Human experience is inherently multisensory, and we use all our senses when we are driving, yet RO's will be limited in their ability to gain authentic sensory information as they will not be 'there'. This suggests that, for remote operators to build up a mental model of the environment that they do not physically occupy, they will need a rich array of sensory information from a variety of different perspectives to help them. The current study aimed to shed some light on the question of what information might be necessary for remote operators to build remote SA quickly and accurately by examining the effect on remote SA of two information sources presented in addition to the standard front camera feed – a rear-view feed and an audio feed.

1.1 Research relating to rear view information and situation awareness

The Highway Code[1] dictates that mirrors should be used before commencing all driving manoeuvres, for example, at roundabouts and junctions and when turning left or right, using the method, "mirror, signal, manoeuvre." According to Regulation 33 of The Road Vehicles (Construction & Use) Regulations 1986[2], all UK vehicles must be fitted with either an internal rear-view mirror or an external mirror if the internal mirror cannot provide an adequate view to the rear (in the case of a panel van). This requirement includes autonomous vehicles. For example, the Tesla Vision model which has a camera located above the rear-view mirror to supplement the eight external cameras (Hawkins, 2019). One would assume that rear-view information would be essential if a human driver had to take over control of an AV in an edge case and build SA of the remote scene. However, in early 2020, AV company Nuro was granted temporary approval in the US for an autonomous

---

[1] https://www.highwaycodeuk.co.uk/using-the-road.html

[2] https://www.legislation.gov.uk/uksi/1986/1078/contents/made

delivery vehicle with no physical side or rear mirrors (Congressional Research Services, 2021). This move suggests that new automation technology is becoming more acceptable, illustrating the current importance of research to examine the information that remote operators might require – introducing new AV models without empirical consideration as to whether the absence of focused rear-view information will have a negative effect on an RO's SA could be risky.

Several studies have specified that rear-view awareness is critical when building up SA in human-robot interaction, particularly in the field of automated vehicles, for example in teleoperated domestic robots (Adamides, Christou, Katsanos, Xenos, & Hadzilacos, 2015), urban search and rescue robots (Drury, Scholtz, & Yanco, 2003; Drury, Keyes, & Yanco, 2007) and automated driving systems (Saffarian, de Winter, & Happee, 2012). Early field trials of semi-autonomous vehicles found that the first thing operators did when the AV encountered a problem was use the cameras to investigate the scene around them, even though this added a fixed 18 seconds of automated panning before any driving could recommence (Burke, Murphy, Rogers, Lumelsky, & Scholtz, 2004). This demonstrates, unsurprisingly, that operators are likely to seek information from the forward, side and rear-view visual fields visual field as a priority when acquiring understanding of the scene and successful control of a vehicle.

Driving requires attentional shifts between the road ahead and the rear and side mirrors in order to track traffic around the vehicle using spatial working memory (Gugerty, 1997). Studies which have measured direct usage of rear-view mirrors and side mirrors (for example, Bolstad, Cuevas, Wang-Costello, Endsley, & Angell, 2010; Lorenz, Kerschbaum, & Schumann, 2014) conclude that participants rarely use the rear-view mirror, despite its purpose as "a display to support SA" (Bolstad et al., 2010, p. 10). Eye tracking studies using driving videos, which have measured the time taken to gain SA after take-over requests (TORs) in AVs, report that participants only distribute equal attention to the road ahead and the mirrors during the first 2-4 seconds, then glance frequency to the rear-view mirror decreases over time (Lu, Coster, & de Winter, 2017). However, highly automated vehicles have the potential to provide enhanced rear-view vision compared to non-automated vehicles, and in remote operations ROs may need to make more use of the rear-view footage that is available to them to immerse themselves in the viewing scene (Saffarian, de Winter, & Happee, 2012). Determining whether operators draw on rear view information to build and maintain SA is thus an important subject of enquiry, and one of the questions addressed by the current research.

1.2 Research relating to audio feeds and situation awareness

The question of whether an auditory feed from the scene can support remote SA is also addressed in the current work, however this question has received relatively little research to date. Nevertheless, the potential importance of including audio from the scene might be argued to be reflected in the decision of many driving researchers to mute audio feeds in their experiments, or instead to play standard highway sounds unrelated to the traffic depicted in stimuli, due to the concern that more realistic sounds from the driver's vehicle and other vehicles, such as the sound of the car engine or tires on the road, may provide useful information about the driving task and therefore risk confounding the manipulations of interest (Lu et al., 2017). According to this view, providing auditory information may help ROs in their formation of remote SA by directing their attention more quickly to salient information in the scene.

Indeed, hearing a relevant sound has been found to direct attention to that sound's location thus enhancing visual perception (Störmer, 2019). In line with this claim, spatially directed auditory warning signals have been used in simulated driving tasks as a means of directing the participants' visual attention to the relevant direction. Ho & Spence (2005) investigated the use of spatial auditory cues in potential emergency driving situations (a car approaching at speed from behind) seen via the rear-view mirror. When participants heard a car horn from the same direction (i.e., the rear) they reacted quicker, by correctly looking in their rear-view mirror to locate the threat, than when the auditory warning signal came from an incorrect direction, for example the front. Further evidence from cross-modal attentional research indicates that multisensory cues are processed more effectively if they are co-located from the same direction, i.e., audio and visual stimuli should both originate either from the front or from the rear (Spence, 2010). This suggests that sensory signals presented from the same direction have the potential to enhance both attention and visual perception, so auditory feeds from around the remote vehicle could support the development and maintenance of ROs remote SA.

One possibility is that remote SA will be improved by the provision of auditory information because the additional input will help operators to feel more 'present' in the remote scene. Presence refers to a subjective perception of being somewhere else, which for an RO, would be a feeling of being inside the remote vehicle through their sensory experience (Georg et al., 2020). Larsson, Västfjäll, Olsson, & Kleiner (2007) investigated congruent and incongruent sounds of a church organ in a virtual environment compared to no sounds, finding that participants rated their sense of presence higher in the sound conditions compared to the no sound conditions. This finding is consistently supported in presence research. A meta-analysis of 83 studies researching presence reported that those which manipulated the relative presence or absence of sound observed that auditory information increased the sense of presence or a sense of "being there" (Cummings & Bailenson, 2016, p. 5).

Audition has been controversially called a “*secondary sense*" compared to vision, despite findings that an absence of auditory information makes our connection with the environment seem weaker or muffled (Larsson et al., 2007, p. 1). Aerospace applications have long made use of providing auditory information to assist pilots coming back into the loop in unmanned aerial vehicles (which require a strong sense of presence to operate them safely). It would be reasonable therefore to assume that providing auditory information would give ROs a greater sense of presence in the remote scene, enhancing remote SA.

1.3 Effect on workload of providing supplementary information to the forward-facing camera

Each additional piece of information provided to an RO is likely to carry a processing burden, in requiring the operator to absorb the information and decide how to act. Strayer and Fisher (2016) showed that as drivers’ involvement in secondary tasks increases, their attention to side and rear-view mirrors dropped. Heenan, Herdman, Brown, & Robert (2014) found that asking participants to hold mobile phone conversations whilst driving had a detrimental effect on their awareness of vehicles to the rear. Gugerty (1997) argued that driving requires attentional shifts between the road ahead and the rear and side mirrors to track traffic around the vehicle using spatial working memory so supplementary auditory information may, in fact, diminish rear-view awareness. This research illustrates that having both visual and auditory sources of sensory input may be distracting and represent a processing burden for ROs.

However, Desai et al (2013) found an increase in perceived workload in conditions where participants achieved lower SA, perhaps because participants had to work harder to build and maintain SA under these conditions. This suggests that, in the current study, conditions which provide rear-view and auditory information may in fact be less demanding than others, despite providing additional information, if they successfully support enhanced SA as we expect.

A careful consideration of the ways in which workload interacts with SA is therefore essential in ensuring the safety of remote intervention. This would ideally disentangle the different types of workload which are assessed by standard measures such as the NASA Task Load Index (NASA- TLX), for example the mental, physical and temporal demands placed on the operator by the task and their resultant effects on performance, effort and frustration (Hart, 2006).

1.4 Confidence in SA

An RO needs to have insight into their own SA levels from moment to moment, such that that they are appropriately confident when their SA is high but also able to acknowledge uncertainty at times when their SA is lower; being confident but wrong is dangerous. This type of subjective awareness has been termed 'meta awareness' or, along signal detection theory paradigms, a ‘Type 2 sensitivity' which refers to whether participants can discriminate between their own correct and incorrect judgements (Fleming & Lau, 2014; McGuinness, 2004). However, a signal detection task is not appropriate to measure SA, unless it has been set up as a clear-cut hazard perception test, as there is no true 'stimulus absent'. Furthermore, the task of gaining SA in driving is unlike the simplistic stimulus detection tasks usually employed by signal detection or ROC studies as remote driving videos will have a myriad of potential stimuli that participants could attend to (Beckstead, 2014; Tekin & Roediger, 2017). Finally, D prime also assumes all task difficulty or stimulus strength is held constant which also cannot be true when gaining and maintaining SA. Despite these issues, metacognitive sensitivity has still been found to be related to SA performance.

To respond to these challenges, there have been attempts to combine both an objective assessment of SA (the accuracy of responses to SA queries), along with a subjective rating of how confident the person is in their answers, for example, allocating either binary scores for high or low confidence, or ranking confidence on a scale (McGuinness, 2004). If participants possess good metacognitive sensitivity, then when they are confident in their SA judgements they are likely to be correct (Fleming & Lau, 2014). The importance of these considerations was demonstrated in a study which assessed fighter pilots' confidence in their SA performance and found that over confidence bias was a significant predictor of worse pilot performance, resulting in lower mission survival on a simulated flight task (Sulistyawati, Wickens, & Chui, 2011). Accordingly, determining whether participants are sensitive to the accuracy of their judgements is a key part of this research.

1.5 Aims of the study

This study uses the validated SA measure reported in Paper 3 to investigate whether the presence of additional sources of visual and audio information improve SA. We presented 16 driving videos from the perspective of the visual feed of a remote operator to get a sense of the issues that will be involved when attempting to build awareness second-hand from a remote scene. We measured differences in participant remote SA, sense of presence and perceived workload in a two-by-two repeated measures design with the factors of rear-view feed (present vs. absent) and audio feed (on vs. off).

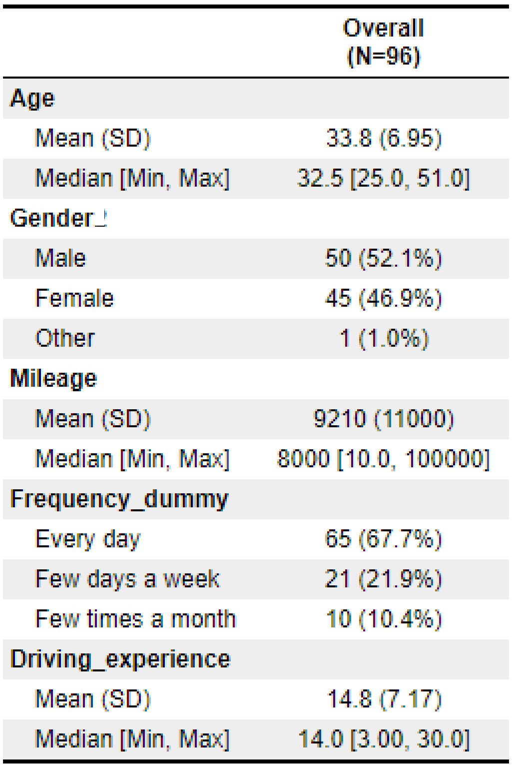
We hypothesized that SA performance would be highest in the condition in which both the rear-view and audio feeds were presented, as this would provide the most information to assist in building awareness. We expected that (a) there would be an effect of audio/no audio and b) there would be an effect of rear-view present/rear-view absent. We also expected participants' sense of presence to be highest in audio-present conditions because more sensory information is presented from the scene.

2.0 Method

2.1 Participants

We used Gorilla (www.gorilla.sc), an online cloud software platform, to collect data for our study. Participants were required to take part using a desktop computer. The videos ranged in size from 3.12 Mb to 19.15Mb, recorded at 60fps and presented on an 811 x 456 screen placement holder. The study took around thirty minutes to complete and participants were compensated £10 an hour for taking part. Participants were informed of the aims of the study and advised that, as with any video watched online, the videos would be temporarily cached in the participant's download file until they had watched them. We informed participants that closing the browser window would terminate their involvement in the study which they could do at any time, permanently deleting their data. All procedures were reviewed and approved by the Royal Holloway Research Ethics Committee. Participants completed a demographic questionnaire about their average mileage and how frequently they drove.

Table 1 Demographic details for n = 96



	Overall (N=96)
Age	
Mean (SD)	33.8 (6.95)
Median [Min, Max]	32.5 [25.0, 51.0]
Gender_	
Male	50 (52.1%)
Female	45 (46.9%)
Other	1 (1.0%)
Mileage	
Mean (SD)	9210 (11000)
Median [Min, Max]	8000 [10.0, 100000]
Frequency_dummy	
Every day	65 (67.7%)
Few days a week	21 (21.9%)
Few times a month	10 (10.4%)
Driving_experience	
Mean (SD)	14.8 (7.17)
Median [Min, Max]	14.0 [3.00, 30.0]

Because the current study used a newly developed methodology (see Paper 3 for details of design and validation), we used the first 24 participants to run a formal power analysis. This indicated that a total of 96 participants were needed for the main study (alpha = 0.05, power = 0.8, Cohen’s f effect size = 0.33) enabling us to detect a small/medium effect size (or relatively small changes in situational awareness). The data from these initial 24 participants were included in the final study as there were no changes to the method or procedure. All participants were UK residents with English as their first language and possessed a full UK driving license for a minimum of three years. Table 1 gives demographic details relating to the frequency of driving, number of years they had held a license and their gender.

2.2 Materials

2.2.1 Creation of driving videos with rear-view present/absent and audio feed on/off

We used 16 driving videos (see Paper 3 for details of their design and creation) demonstrating a range of road types with varying speed limits between 30-70 mph such as motorways, A and B roads, rural roads and residential areas and included a range of driving conditions including fog, low sun and heavy traffic. Video and audio recordings of the forward and rear views were filmed using two GoPro6 cameras mounted on a 2016 petrol Dacai Duster 4x4 using the GoPro Suction Cup Mount. Two lenses captured the video with a spherical field of view (FOV) of 360 degrees (sensor size 1/2.3", with aperture of f/2.8).

We used the audio feed from the bonnet-mounted GoPro camera to provide the audio track for the videos that were presented with sound. The volume levelling of each video was lowered to 88% of the original sound bar volume level, to reduce the effect of potential distortions resulting from differences in the quality of participants' loudspeakers. Participants were instructed to use speakers, not headphones, for listening to the experimental audio. They were played an audio recording which asked them to turn on their speakers and adjust them to a comfortable level to hear the audio track, then completed an attention check which asked them what they had had for their dinner that evening to show they had listened to the recording. The picture and audio were faded in at the start and end of each video to provide a smooth entrance and exit of the stimuli and to avoid any jarring effects of the sudden introduction of sound in the audio on conditions.

In the rear-view present videos, the rear-view footage was presented to the left of the forward-facing imaging and superimposed onto the forward-facing scene. All sixteen separate videos were checked to ensure information such as road signs were not occluded by the rear-view video. As the rear-view footage had been filmed attached to the rear windscreen pointing outwards, the images for the 16 driving videos were rotated 180^0 to create the impression of reflecting the road behind the car. Thus, any cars passing the 'driver' in the forward-facing image appeared on the right side of the road in the rear-view footage. The size of the rear-view footage was set at a ratio of 30.2 % of the size of the forward-facing video, based on the usual proportional size of a rear-view mirror to a standard car windshield.

Four versions of each video were created (shown in in Table 2) by either including or excluding the rear-view footage and the audio feed, giving 64 distinct videos in total. Videos were presented in blocks of 4 in each condition.

Table 2 Description of the experimental conditions with abbreviations

Condition label	Description
No audio, no rear-view (NANR)	Forward facing video only with *no audio*
Audio, no rear-view (ANR)	Forward facing video only *with* audio
Rear-view, no audio (RNA)	Rear and forward-facing views with *no audio*
Rear-view, audio (RA)	Rear and forward-facing views *with* audio

Figure 1 illustrates the visual differences between rear-view present and rear-view absent conditions.

Figure 1 Example of the rear-view absent presentation of the driving video (left) and rear-view present images (right). The rear-view image is superimposed over the forward-facing view in the top left corner of the screen in rear-view present videos.

2.2.2 SA questions for the 16 driving videos

We used the SA questions matched to each video, which were developed and validated in Paper 3. SA questions could be answered in each condition (i.e. there were no questions related to information only visible from the rear view feed which would be unanswerable in the rear-view absent conditions and no questions relating to the audio feed which would be unanswerable in the audio-absent conditions). Each question was designed to measure a particular level of SA (perception, comprehension, and prediction) and there were an equal number of questions presented for each SA level.

Perception questions asked about visual information present in the remote scene in the driving video for example "What colour were the traffic lights?". Comprehension questions were either from the perspective of the driver or other road users and measured their understanding of the remote scene, for example "Why are you waiting behind the blue car?". Prediction questions tested participants' ability to use information in the remote scene to make future predictions about what may happen in the next few seconds based on the ending of the video, for example, "What direction will you go in the next few seconds?". Participants wrote their answer to the SA question in a text box on the screen which had no restrictions on text response. All questions required a written response before the participant could move on to the next video in the block. The order of SA type of questions varied randomly to prevent order effects (for example, if perception questions were always asked first then participants may learn to look for these details first in the driving videos and/or these questions could be more susceptible to forgetting effects than the questions asked last).

2.3 Design

The study employed a 2x2 within-subjects factorial design: Audio (on, off) x Rear-view (present, absent). We presented sixteen videos each of which had 4 possible versions corresponding to the 4 conditions; rear with audio (RA), rear no audio (RNA), no audio no rear (NANR), and audio no rear (ANR). Condition was blocked, such that participants saw all four videos for each condition as part of a single block, and the block order was counterbalanced across participants.

The allocation of videos to blocks was randomised for each participant. All participants saw 16 videos in total, divided into the 4 blocks, but within each block, participants saw a different set of videos which were allocated at random with no replacement. For example, Video #1 may be RA presentation for Participant 1, but Participant 2 may view Video #1 as the NANR version. All versions of all videos were sampled equally across the 96 participants.

Each block was preceded by a white screen with writing which informed the participant which condition they were about to see for example *"Audio Rear. You are now about to see four videos with audio and rear-view presentation. Make sure your sound is switched on".* This was to ensure that conditions which required audio were flagged to the participants in case of any problems with sound, which could be reported in the debrief.

2.4 Procedure

Participants watched 16 videos of driving scenes with four videos in each condition, presented in a blocked design. At the end of each video, participants were asked three separate questions designed to measure each level of SA (perception, comprehension, and prediction – although recall that the order of these question types was randomised) and were asked to enter their answer to each question in a text box.

After each SA response, participants were asked '*How confident are you in your answer'*? This was intended to measure if participants knew whether their answers reflected correct information or merely guess work. Participants indicated their confidence judgement on a four-point Likert scale as this has been argued to be the optimal number of response options in relation to survey measurement of cognitive properties (Beckstead, 2014). Options ranged from 1 meaning "not at all confident or guessing" to 4 which was "completely confident". In the pilot study (reported in Paper 3), participants had made comments about the speed the car was travelling in, so we added a speed question after each of the videos, which asked "*how fast was the car travelling (in mph) in this video*?". Technical problems with the interpretation of the responses to this question meant that this data was not analysed further although we recognise that this is still an interesting area of study.

To gauge whether some of the experimental conditions were more demanding than others, at the end of each block, participants were asked a workload question, based on one sub-section of the TLX-R relating to how much mental activity was required by the task, namely "*Was answering the questions on the videos in this section easy or demanding, simple or complex?"* (Hart, 2006). Participants indicated on a sliding scale from 'very low' on the left to 'very high' on the right side of the slider. Workload scores were converted to a score out of 100 where 1 was 'low demand' and 100 was 'high demand'.

In order to assess 'presence' as a potential contributing factor to any finding of enhanced SA, we added a question to the procedure in the main experiment to measure participants' sense of engagement in the driving videos under each condition; "*Please indicate on the scale how immersed you were in this section*". They answered on a sliding scale from 'completely immersed' (left side of the scale) to 'not at all engaged' (right side of the scale). The question was reverse coded, meaning that high sense of presence was set at the opposite extreme to the high workload question response. This was intended to encourage participants to pay attention to the verbal indicators and reduce any influence of the previous question. Presence scores were coded out of 100 where 1 corresponded to being *'completely immersed'* and 100 was *'not at all engaged'*.

Participants were debriefed and asked if they had any comments. Further qualitative data was collected by the questions, *"Did you find the audio helpful in the videos that you watched which had sound?"* and " *Did you find the rear-view helpful when it was present in the videos you watched?"* although participants were permitted to leave these questions unanswered if they wished. Few participants responded to these questions, so this data was not analysed further. Figure 2 shows the complete procedure for the main study.

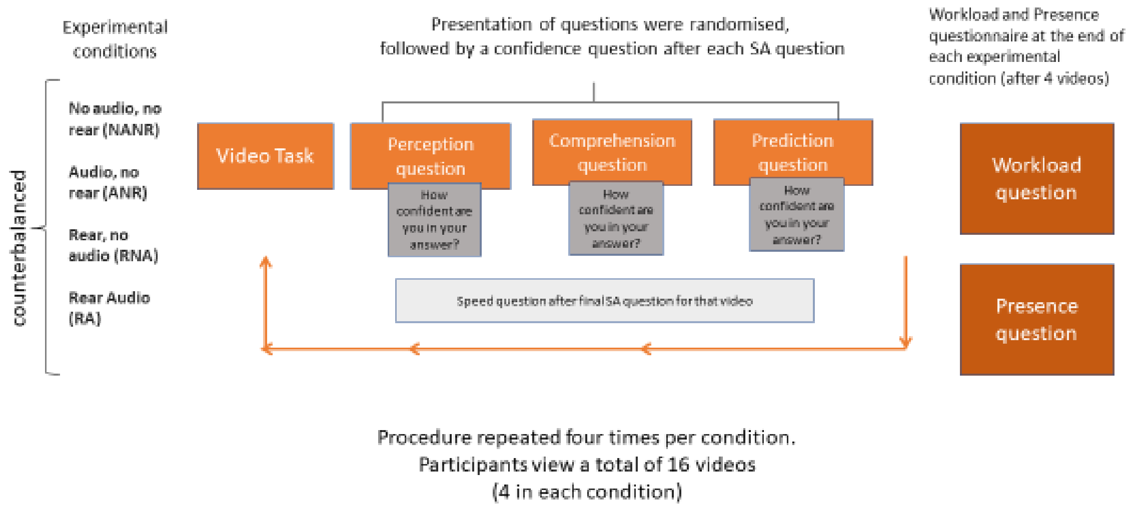


Figure 2 Procedure of the experiment.

3. Results

3.1 Dependent measures

Each video had three SA questions individually designed based on the events of that video. Correct answers were agreed by the research team from viewing the videos and participant responses were coded blind by the lead researcher. We conducted three separate analyses on the SA scores, including a novel analysis that incorporated confidence judgements into remote SA scoring. We also conducted analyses on the responses to the speed question after each video and the presence and workload questions after the end of each block.

3.2 Descriptive statistics

Each participant saw 4 videos in each condition with 3 questions on each level of SA asked for each video. One participant could not answer the question " Why should you change down to a lower gear in the next few seconds?" as they drove an automatic vehicle. Another participant had a loading error on the Gorilla platform for one question so could not provide an answer. Rather than score both events as incorrect, penalising their score, we calculated their proportional performance score out of 11 instead of the possible 12. No outliers were removed. Two participants were excluded as they did not meet the sample restrictions.

Formal observations on SA performance score for each level (Perception, Comprehension, Prediction) were not carried out due to the low number of observations (4) for each participant in

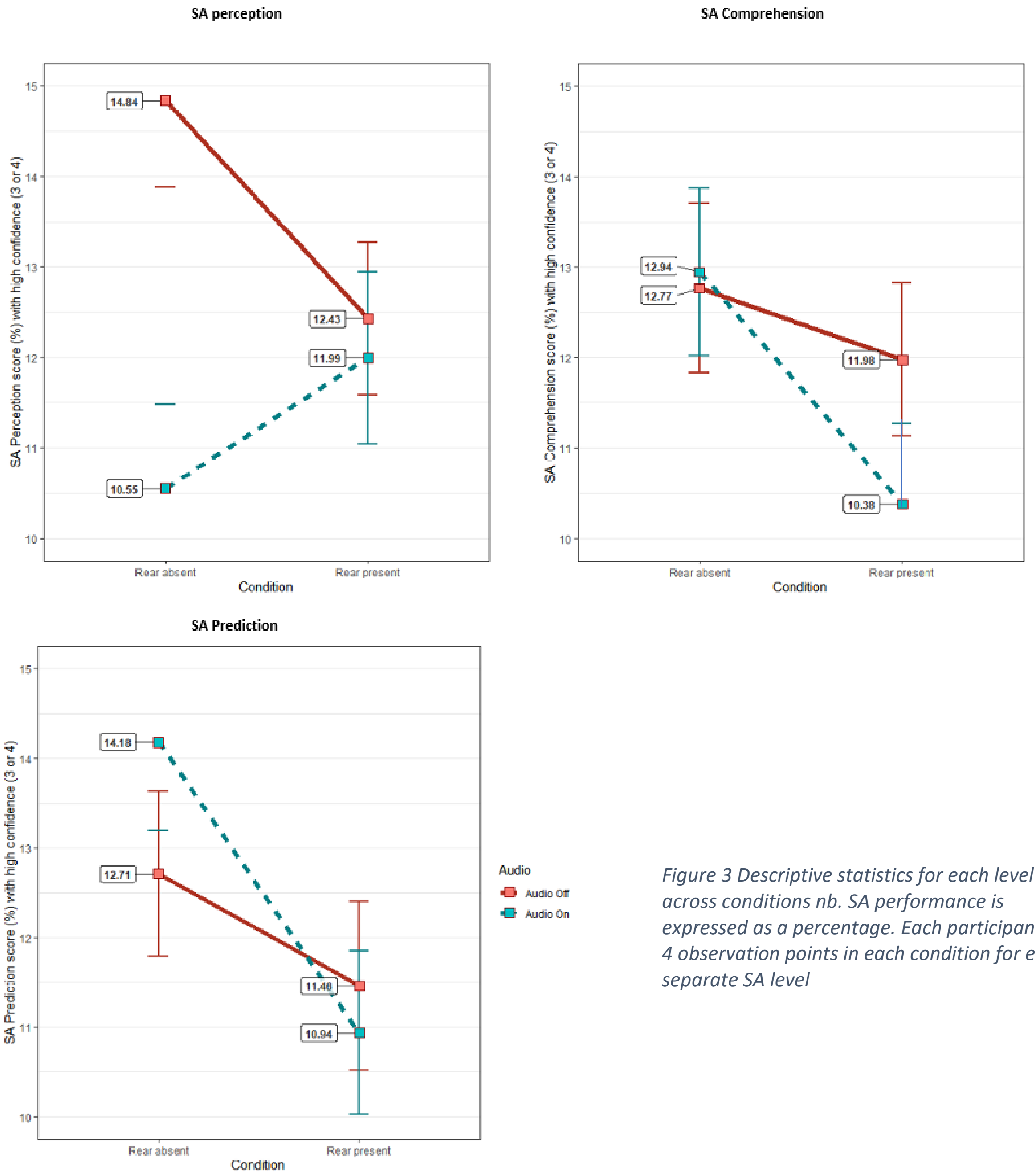


Figure 3 Descriptive statistics for each level of SA across conditions nb. SA performance is expressed as a percentage. Each participant had 4 observation points in each condition for each separate SA level

each condition, however Figure 3 shows visual comparisons (in %) of overall SA performance across each condition for reference.

3.3 Analysis #1: SA Performance Score

Participant answers were given a binary score of 1 or 0 depending on whether they matched the correct answer determined by the research team from viewing the video. Responses for perception, comprehension and prediction were summed to derive an overall SA performance measure with the maximum score in each condition being 12, giving a total possible score of 48 summed overall. Analysis 1 quantified SA as the total percentage of correct responses in each condition (RA, RNA, ANR, NANR).

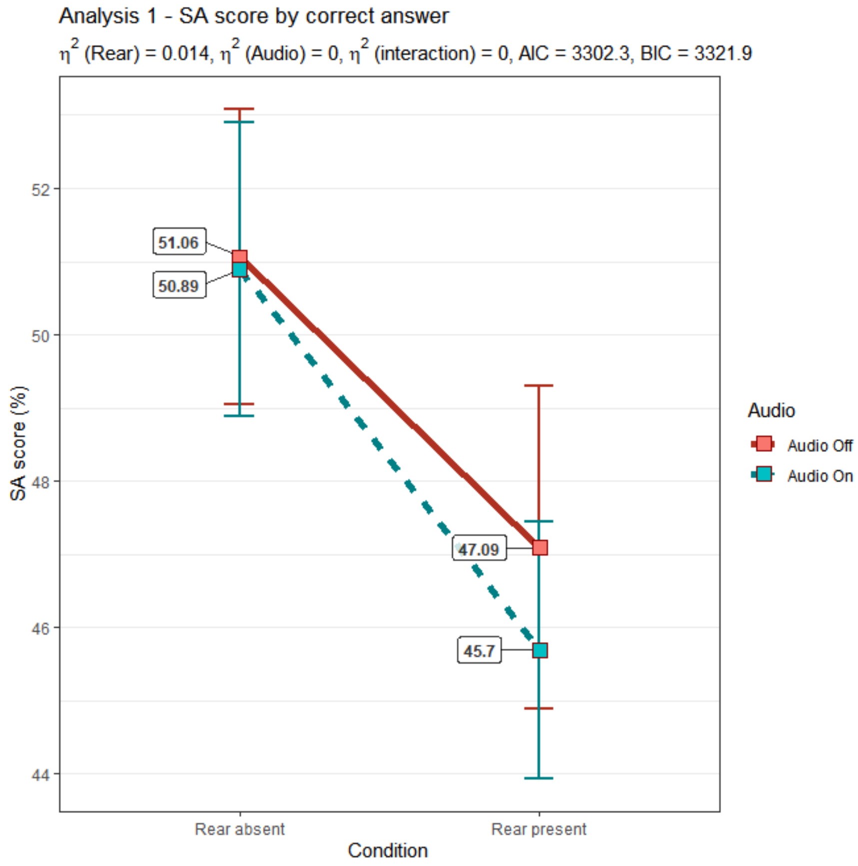


Figure 4 Analysis #1 - Plot of main effects of audio and rear n= 94

Analysis 1 showed that participant SA performance was, overall, fairly poor in all conditions, with around 50% mean accuracy, although this was not due to chance; participants gave a correct answer out of many different possible answers to each question. A 2x2 repeated measures ANOVA was carried out with factors of rear-view feed (present, absent) and audio feed (on, off). We found a significant main effect of rear-view feed (F(1,93) = 10.70, $p < 0.02$, np = 0.01) with lower scores when the feed was present than when it was absent (see Figure 4). There was no main effect of audio feed ($F < 1$) and no interaction between the two factors ($F<1$).

3.4 Analysis #2: SA performance with high confidence

SA involves **knowing** what is going on around you, so it is important to attempt to identify instances of luck or guess work when assessing SA performance. Analysis 2 adopted a novel measure of combining performance accuracy and high confidence, such that answers were only scored as correct if they were given with high confidence (selected 4 or 3 on the confidence scale) (see Table 3).

Table 3 Analysis #2 High confidence and correct answer constitutes a hit

Decision	High confidence (4/3)	Low confidence (2/1)
Correct	1	0
Incorrect	0	0

Performance was quantified as the total percentage (out of a total of 12) of correct, high confidence responses in each condition (RA, RNA, ANR, NANR).

Unsurprisingly, this reduced the scores overall by more than 25% (shown in Figure 5). However, in line with the results of Analysis 1, the 2x2 repeated measures ANOVA with the factors of rear-view feed (present, absent) and audio feed (on, off) identified a significant main effect of rear-view feed (F(1,93) = 11.57, p < 0.01, np = 0.01) with no main effect of audio feed (F (1, 93) = 2.63, p = 0.11) and no interaction (F<1).

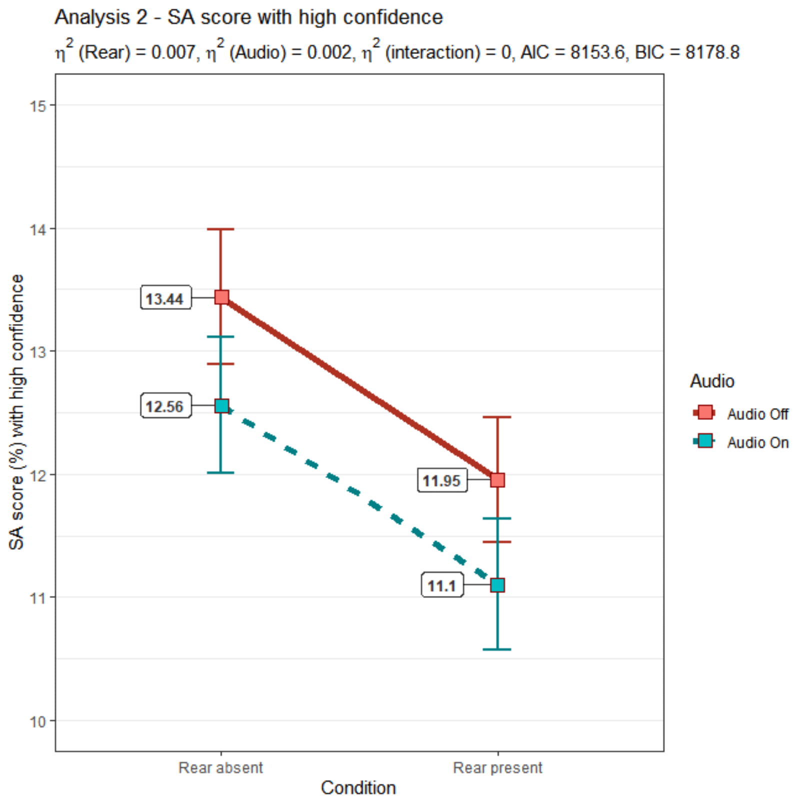


Figure 5 Analysis #2 - Plot of SA performance with high confidence for audio and rear conditions (n= 94)

3.5 Analysis 3: Correct SA with high confidence separated by SA level

Having carefully designed different SA questions for each level of SA based on the content of each video, we further analysed SA performance across perception, comprehension, and prediction questions. We looked separately at the results from the three SA levels collapsed across conditions to compare performance across SA levels. Visual inspection of descriptive statistics revealed variance in performance with all means around 12% (see Figure 6).

A one-way repeated measures ANOVA was performed to compare the effect of SA level on SA performance. Results revealed that there was no statistically significant difference in SA performance between the three groups (F(1, 93) = 0.295, p = 0.745).

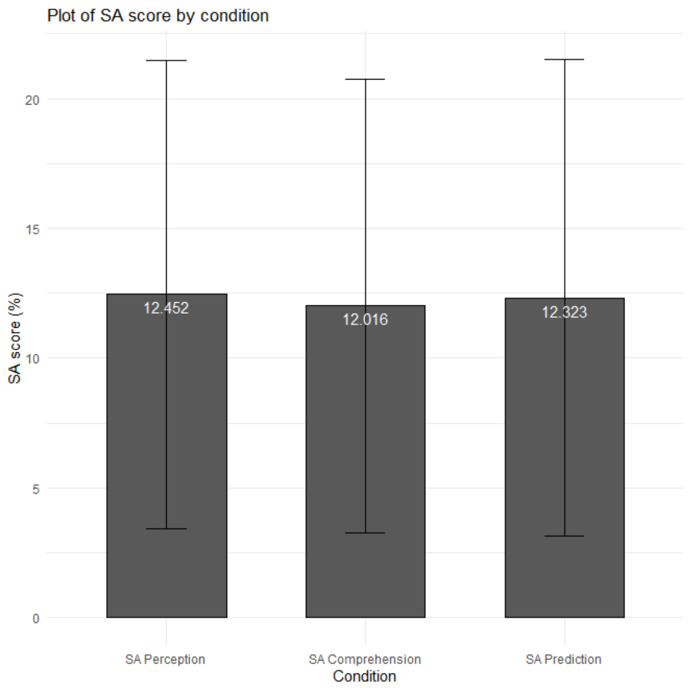


Figure 6 Bar plot of mean SA performance score (%) across SA Level

3.6 Sense of presence across conditions

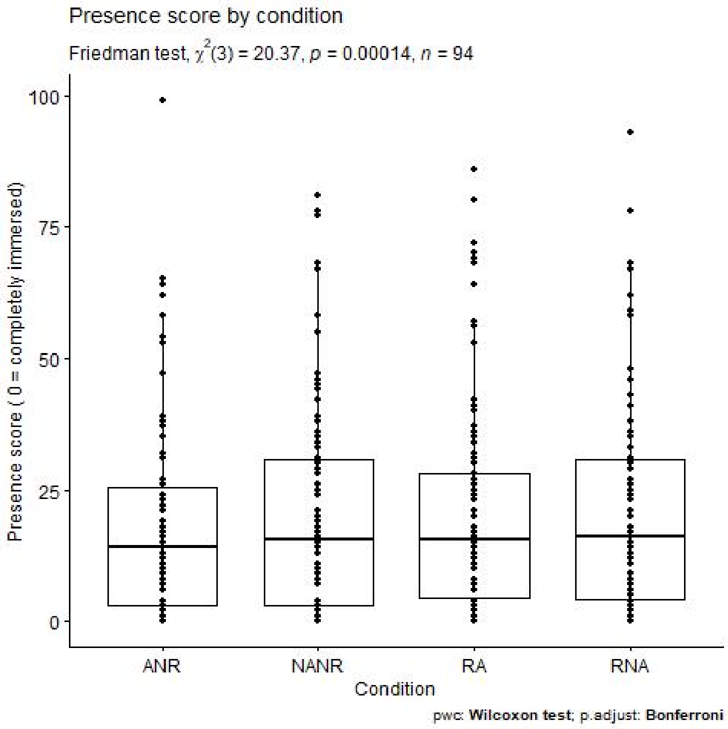


Figure 7 Plot of presence scores across condition with Friedman test

Participants gave individual scores from 0-100 for each condition of the study, based on their assessment of how much of a sense of presence the videos in that condition provided. A score of 0 meant they were completely immersed but this was counter intuitive so, for clarity, we converted the scoring so that a score of 100 reflected complete immersion. Table 4 shows that means for all four conditions were below 22, suggesting that the participants did not feel a strong sense of presence in the remote driving scenes.

Data were not normally distributed and the small number of observations (1) per participant meant that the assumptions of ANOVA were not met. Presence data was analysed using a non-parametric Friedman test which indicated that presence scores did differ significantly between the four conditions, $X^2_F(3) = 20.4$, p = 0.0001, $\eta^2_g = 0.0722$.

Table 4 Summary statistics for presence question across conditions

Rear	Present		Absent		Present	Absent	Present	Absent
Audio	On	Off	On	Off	On		Off	
M	20.9	21.5	18.3	20.1	20.9	18.3	21.5	20.1
sd	21.0	20.4	19.2	20.7	21.0	19.2	20.4	20.7

We followed up by conducting a pairwise Wilcoxon signed-rank tests for identifying which groups are different. Pairwise Wilcoxon signed rank tests between 4 conditions with Bonferroni corrections (significance level adjusted from α =.05 to α = .0125), revealed statistically significant differences in presence levels only between ANR and RNA (p = 0.0005). Videos viewed with audio feeds without rear-view information (M^{ANR}= 18.3) resulted in a lower sense of presence rating with the highest sense of presence in the conditions which had rear-view information present and audio feed off (M^{RNA} = 21.5).

3.7 Workload measure across conditions

As the study was hosted online, the current study only included a brief assessment of SA workload by including the mental demand subscale of the NASA-TLX, "*Was answering the questions on the videos in this section easy or demanding, simple or complex*?". Workload responses were scored from 0-100 for each condition of the study (100 = very high). Workload scores were analysed using a non-parametric test as the assumptions of ANOVA were not met due to small sample size and non-normally distributed data.

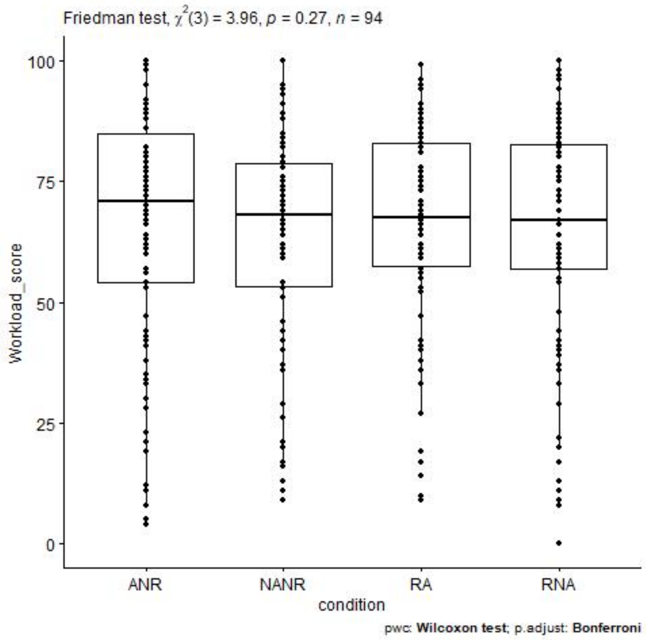


Figure 8 NASA-TLX Mental demand across condition

While descriptive statistics revealed that the NANR condition was rated the lowest workload (M^{NANR} = 63.9) compared to the other conditions (see Table 5), a non-parametric Friedman test used to assess whether there were differences between the groups revealed that this difference was not significant (X^2_F (3) = 3.96, p = 0.27).

Table 5 Workload scores (n=94) nb. 100 = very high mental demand

Rear	Present		Absent		Present	Absent	Present	Absent
Audio	On	Off	On	Off	On		Off	
M	65.9	64.5	65.2	63.9	65.9	65.2	64.5	63.9
sd	21.6	24.4	24.6	22.3	21.6	24.6	24.4	22.3

4. Discussion

4.1 Comparison with previous research

We investigated the effect of the provision of rear-view and audio information on SA for remote driving scenes and consistently found worse SA performance in the presence (vs. absence) of the rear-view feed. Research findings in this study at first glance appear to concur with previous studies (Bolstad et al., 2010; Gugerty, 1997; Lorenz et al., 2014) which found that participants rarely use the rear-view mirror and that secondary tasks may diminish attention to rear-view mirrors, but in fact our findings suggest that participants found the presence of a rear-view feed distracting rather than ignoring it. Conceivably, the extra visual detail provided in the rear-view present conditions functioned less as an aid to SA and more as a distraction, such that participants were able to concentrate better and pick up more information when only the forward-facing camera was present. However, there is a possibility that the length of the videos may have reduced the efficacy of the rear-view footage. Previous research (Lu et al., 2017) indicated that attention to the rear-view mirror decreased over time, so if we had frozen the video earlier (or at a point in the video when an event was occurring for which the rear-view information may have been useful) we may have found more of an effect of people incorporating it into their SA. Furthermore, when driving in real life, moving your head slightly allows you to see around the mirror, whereas in our study, the presentation of the rear-view feed permanently occluded a percentage of the screen, which was visible in the rear view absent condition. Although the SA questions were carefully devised to ensure that the top left corner of the screen contained no pertinent information, in future studies, it would be prudent to cover the same area with a black box in the rear view absent condition for parity.

We found no effect of audio feed on SA performance, which implies that muting audio feeds to reduce confounding effects in traffic studies as in Lu et al. (2017), may not always be necessary. However, none of the videos in this study had SA questions which could have been directly assisted by sound cues which may not be the case in other studies. One of the videos in our study had the sound of a car horn which, according to research by Ho & Spence (2005), could have helped with SA by directing attention towards the relevant visual information. SA scores for audio conditions for this video were higher than the non-audio conditions, which may provide some anecdotal indication that the salient sound had a positive effect on people’s ability to extract correct information from the scene. This might be an interesting area for future research, particularly in relation to the use of 3D audio which might provide more effective spatial cuing effects. However, we also found no evidence that the audio feed affected participant's sense of presence which contradicts consistent findings that participants rated their sense of presence higher in conditions which provided sound (Cummings & Bailenson, 2016).

We found the highest sense of presence in the conditions which had rear-view information present but overall, the scores were all low in terms how much of a sense of presence the driving videos produced. Potentially, this could be due to the small number of data points in our data collection (only one for each participant in each condition) which may mean that our study was not sufficiently powered in this question to detect an effect. Also, the presence question was reverse coded, meaning that high sense of presence was set at the opposite extreme to the high workload question response. This was intended to encourage participants to pay attention to the verbal indicators and reduce any influence of the previous question but, possibly, participants had misinterpreted the scale.

Previous research (Desai et al, 2013) found an increase in perceived workload in conditions where participants achieved lower SA, so, in line with our findings, we would have expected to find lower workload scores in the rear-view conditions as they had the lowest SA scores. Instead, we found no effect of workload in this study, but this may be because we only used one sub-section of the NASA-TLX scale and only one observation per participant in each condition.

The SA questions in our study may have acted as a prompt to short term recall, reminding participants of what had happened in the video or enabling them to make educated deductions based on driving experience. When we limited the measurement of SA performance to correct answers with high confidence it lowered SA scores considerably, showing that many participants may have been guessing to at least some extent (or at least responding with low confidence) throughout the study. Previous research found that overconfidence bias was higher for Level 3 SA (projections) than for Level 1 or 2 SA whereas we observed this pattern to be consistent across all SA levels (Sulistyawati et al., 2009). Given these weaknesses in current practice, it may be necessary to reconsider approaches to measuring driving SA during remote operation.

The exploratory factor analysis reported in Paper 3 on the 48 items from the SA measure used in this study found four underlying constructs (‘spatial and environmental awareness’, ‘anticipatory hazards’, ‘dynamic driving actions’ and ‘other road users’) and indicated that separate SA questions on the three ‘levels’ of SA (perception, comprehension and prediction) were combined when building awareness of a remote driving scene. In the current study, we found no significant differences in SA performance across the three SA levels: although the SA questions were tested on a relatively small sample, this finding offers further support for the theory that holding them separate may not be the most appropriate way to assess SA in remote operation. Whilst undoubtedly still in the early stages of development, future studies could design questions to measure driving SA in accordance with the four underlying constructs established in Paper 3, rather than the conventional three ‘levels’, to determine if this approach represents a more appropriate behavioural measure to understand driving SA.

4.2 Limitations of research

Although this study does advance our understanding of the challenges confronting remote operators attempting to gain SA from remote situations using 2D modes of presentation, there are clear limitations in drawing conclusions from the data. When driving a car, attention is usually diverted by other driving tasks such as changing gears, depressing pedals and other peripheral distractions (e.g., changing music). In this task, participants were not actively driving and so were not subject to these distractions, so the results cannot be taken to represent authentic driving experiences, although an RO is unlikely to be carrying out these tasks either in use cases such as remote supervision. Gugerty (1997) found that if participants were in control of the driving task, they had better recall of hazards rather than when they were a passenger suggesting that the nature of our task may have limited the SA that participants may have acquired if actively driving. However, Mackenzie & Harris (2015) showed that participants were slower to detect hazards when driving themselves as opposed to passively viewing a video so equally, by this argument, participants’ SA could have been augmented in our study by the remote viewing context. Future studies could ask participants to remotely control a robot in order to achieve a more interactive task. Increased interaction may also prompt more engagement with the rear-view feed and thus could provide more illustration of how an RO would make use of the rear-view information in the context of more active teleoperation.

We presented audio information in our driving videos from the audio captured from the bonnet-mounted GoPro camera, however this meant that wind noise was more audible than for a driver inside the vehicle which may have been distracting for the participants in our study. This may explain why the highest sense of presence was reported in the conditions which had rear-view information present and audio feed off. Guided by previous evidence (Spence, 2010), we could have used the audio from the rear camera instead, providing salient information which could have served to spatially orientate attention to the rear, improving SA performance in the audio on conditions. However, as the study was hosted online, participants were at home using their own speakers so the recording noise would not be spatially accurate when they were participating in the study.

Some manoeuvres, such as entering and exiting a roundabout, require glances to side mirrors and to check the blind spot which couldn't be done in this study. Participants who referred to the rear-view feed in the debrief cited it as helpful during periods of activity such as approaching a roundabout or assessing hazards. On the other hand, videos Rural#1 and Rural#2 depicted roads that were long and straight with minimal traffic, rendering the rear-view information redundant. If we had frozen the video at a point when an event was occurring that may have been necessary to look in the rear-view feed for more information, we may have had more reference to it in the rear-view present conditions. In addition, because of the need to match the SA questions across all four conditions, none of the questions asked about information contained within the rear-view feed, meaning that the presence of the rear-view feed could rationally be discounted as a distraction, and it may in fact arguably represent ‘better’ SA to ignore it rather than attending to it, thus reducing workload. Participants took part in all conditions, so they may have quickly worked out that the information in the rear-view was not necessary to answer the questions after the first few videos. Although we do not endorse designers of interfaces for ROs removing rear-view feeds, the knowledge that, contingent on task goals, providing rear view information is not always essential to building successful SA, will be a useful consideration although this conclusion must remain tentative for now given the limitations outlined above.

4.3 Recommendations for future research and conclusions

Although driving simulator studies regularly include a rear-view mirror as part of the design apparatus (for example Crundall et al., 2012; Eriksson & Stanton, 2016; Gold, Körber, Lechner, & Bengler, 2016; Hergeth, Lorenz, Vilimek, & Krems, 2015; Merat, Jamson, Lai, Daly, & Carsten, 2014; Mok, Johns, Miller, & Ju, 2017; Palazzi, Abati, Calderara, Solera, & Cucchiara, 2019; Van Den Beukel & Van Der Voort, 2013), there is limited behavioural data reported on its usage. Our study represents an important consideration on the use of the rear-view feeds in remote contexts particularly as new camera technologies emerge to replace traditional rear-view and side mirrors as visual aids to driving. It is certain that providing both auditory and visual feed from the vehicles’ location will have value under some circumstances in assisting ROs in their development of SA, but our research highlights that, depending on task goals, these information sources are not always as helpful as we might assume and may occasionally impair performance. These initial findings provide an important first step in addressing this issue, given that little other research has directly measured the effect of rear-view presentation. Further research into employing spatial audio in remote contexts to enhance attention and direct visual perception is also a critical direction for designers of remote interfaces to allow ROs to develop a sense of presence in the remote scene.

References

Adamides, G., Christou, G., Katsanos, C., Xenos, M., & Hadzilacos, T. (2015). Usability guidelines for the design of robot teleoperation: A taxonomy. *IEEE Transactions on Human-Machine Systems*, *45*(2), 256–262. https://doi.org/10.1109/THMS.2014.2371048

Beckstead, J. W. (2014). On measurements and their quality. Paper 4: Verbal anchors and the number of response options in rating scales. *International Journal of Nursing Studies*, *51*(5), 807–814. https://doi.org/10.1016/j.ijnurstu.2013.09.004

Bolstad, C. a, Cuevas, H., Wang-Costello, J., Endsley, M. R., & Angell, L. S. (2010). Measurement of situation awareness for automobile technologies of the future. *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*, *4970*(June), 195–213. https://doi.org/10.4271/R-402

Burke, J. L., Murphy, R. R., Rogers, E., Lumelsky, V. J., & Scholtz, J. (2004). Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *34*(2), 103–112. https://doi.org/10.1109/TSMCC.2004.826287

Congressional Research Services. (2021). *Issues in Autonomous Vehicle Testing and Deployment Issues in Autonomous Vehicle Testing and Deployment*.

Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis and Prevention*, *45*, 600–609. https://doi.org/10.1016/j.aap.2011.09.049

Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Desai, M, Medvedev, M, Vazquez, M, McSheehy, S, Gadea-Omelchenko, S, Bruggeman, C, Steinfeld, A & Yanco, H. (2013). Influence of Situation Awareness on Control Allocation for Remote Robots. *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*. https://doi.org/10.1109/TePRA.2013.6556372

Drury, J.L., Scholtz, J., & Yanco, H. A. (2003). Awareness in human-robot interactions. *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, *1*(October), 912–918. https://doi.org/10.1109/ICSMC.2003.1243931

Drury, Jill L, Keyes, B., & Yanco, H. A. (2007). *LASSOing HRI : Analyzing Situation Awareness in Map-Centric and Video-Centric Interfaces*. 279–286.

Endsley, M R. (2000). Theoretical underpinnings of situation awareness: A critical review. In D. . Endsley, M.R and Garland (Ed.), *Situation awareness analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Aerospace and Electronics Conference*, 789–795. https://doi.org/10.1109/NAECON.1988.195097

Eriksson, A., & Stanton, N. A. (2016). *Takeover Time in Highly Automated Vehicles : Noncritical Transitions to and From Manual Control*. https://doi.org/10.1177/0018720816685832

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*(JULY), 1–9. https://doi.org/10.3389/fnhum.2014.00443

Georg, J. M., Putz, E., & Diermeyer, F. (2020). Longtime Effects of Videoquality, Videocanvases and Displays on Situation Awareness during Teleoperation of Automated Vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *2020-Octob*, 248–255. https://doi.org/10.1109/SMC42975.2020.9283364

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over Control from Highly Automated Vehicles in Complex Traffic Situations. *Human Factors*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Gugerty, L. J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. *Journal of Experimental Psychology: Applied*, *3*(1), 42–66. https://doi.org/10.1037/1076-898X.3.1.42

Hart, S. G. (2006). NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER. *Proceedings of The Human Factors and Ergonomics Society 50th Annual Meeting*, 904–908.

Hawkins, A. J. (2019). It's Elon Musk vs. everyone else in the race for fully driverless cars.

Heenan, A., Herdman, C. M., Brown, M. S., & Robert, N. (2014). Effects of conversation on situation awareness and working memory in simulated driving. *Human Factors*, *56*(6), 1077–1092. https://doi.org/10.1177/0018720813519265

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2015). Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust during Highly Automated Driving. *Human Factors*, *58*(3), 509–519. https://doi.org/10.1177/0018720815625744

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied*, *11*(3), 157–174. https://doi.org/10.1037/1076-898X.11.3.157

Larsson, A. F. L., Kircher, K., & Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 229–237. https://doi.org/10.1016/j.trf.2014.05.008

Larsson, P., Västfjäll, D., Olsson, P., & Kleiner, M. (2007). *When What You Hear is What You See : Presence and Auditory-Visual Integration in Virtual Environments Applied Acoustics , Chalmers University of Technology , Göteborg , Sweden , 2 Department of*. 11–18.

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society*, *2014-Janua*, 1681–1685. https://doi.org/10.1177/1541931214581351

Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, *60*, 293–304. https://doi.org/10.1016/j.apergo.2016.12.003

Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, *23*(6), 736–757. https://doi.org/10.1080/13506285.2015.1079583

McGuinness, B. (2004). Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings. *2004 Command and Control Research and Technology Symposium*, 1–14. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA465817

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Mok, B., Johns, M., Miller, D., & Ju, W. (2017). Tunneled In: Drivers with Active Secondary Tasks Need More Time to Transition from Automation. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 2840–2844. https://doi.org/10.1145/3025453.3025713

Mutzenich, C., Durant, S., Helman, S., & Dalton, P. (2021a). Situation Awareness in Remote Operators of Autonomous Vehicles: Developing a Taxonomy of Situation Awareness in Video-Relays of Driving Scenes. *Frontiers in Psychology*, *12*(November). https://doi.org/10.3389/fpsyg.2021.727500

Mutzenich, C., Durant, S., Helman, S., & Dalton, P. (2021b). Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles. *Cognitive Research: Principles and Implications*, *6*(1). https://doi.org/10.1186/s41235-021-00271-8

Palazzi, A., Abati, D., Calderara, S., Solera, F., & Cucchiara, R. (2019). Predicting the Driver's Focus of Attention: The DR(eye)VE Project. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(7), 1720–1733. https://doi.org/10.1109/TPAMI.2018.2845370

Saffarian, M., de Winter, J. C. F., & Happee, R. (2012). Automated Driving: Human-Factors Issues and Design Solutions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 2296–2300. https://doi.org/10.1177/1071181312561483

Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*, 182–200. https://doi.org/10.1111/j.1749-6632.2010.05440.x

Störmer, V. S. (2019). Orienting spatial attention to sounds enhances visual processing. *Current Opinion in Psychology*, *29*, 193–198. https://doi.org/10.1016/j.copsyc.2019.03.010

Strayer, D. L., & Fisher, D. L. (2016). SPIDER: A Framework for Understanding Driver Distraction. *Human Factors*, *58*(1), 5–12. https://doi.org/10.1177/0018720815619074

Sulistyawati, K., Wickens, C. D., & Chui, Y. P. (2011). Prediction in situation awareness: Confidence bias and underlying cognitiv abilities. *International Journal of Aviation Psychology*, *21*(2), 153–174. https://doi.org/10.1080/10508414.2011.556492

Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications*, *2*(1), 1–13. https://doi.org/10.1186/s41235-017-0086-z

Van Den Beukel, A. P., & Van Der Voort, M. C. (2013). The influence of time-criticality on Situation Awareness when retrieving human control after automated driving. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, (Itsc), 2000–2005. https://doi.org/10.1109/ITSC.2013.6728523

Paper 5

Effect of two formats of presentation (HMD-360 and 360 screen-based) on remote supervision of automated vehicles using a choice decision task

Abstract

Remote operators (ROs) of automated vehicles will be unable to provide remote supervision until they gain necessary situation awareness (SA). This study investigates remote supervision of automated vehicles using a choice decision task to test the effect of two formats of presentation of 360° driving videos: 1) in a head mounted display (HMD-360) where field of view (FOV) changes were based on head movement and 2) screen-based (SB) mouse-controlled FOV. Participants viewed 60 videos, either in an HMD or via screen-based presentation, filmed from the starting perspective of a stationary car depicting likely scenarios which would cause an AV to require human assistance, and decided for each scenario the direction that the vehicle should take next (left, right, continue forward or reverse). Participant decision time was recorded via a keypress and Tobii Pro Lab software together with eye movement data collected using an inbuilt infrared Tobii Pro Lab eye tracker in an HTC VIVE headset (HMD-360 condition) or via Tobii Pro Nano screen-based eye-tracker (SB condition). Results were analysed using a linear mixed effects model and showed, on average, decisions were made in the HMD-360 condition in 80% of the time required in the SB condition with decisions getting slightly quicker for each trial over 30 trials. Participants were more than twice as likely to give correct answers when videos were presented in the HMD-360 condition. Eye tracking analyses showed there were significant differences in the duration of fixations in in areas of interest (defined by dividing the 360° video image into four equal parts) depending on whether it was presented in SB or HMD-360 format. It is recommended that companies offering remote supervision of AVs explore the potential of HMD-360 presentation to enhance operator's SA in the remote scene. We suggest that it would be most appropriate to limit HMD to shorter periods of operation to reduce negative workload issues, for example when playing a supervisory role in fleet operations.

1.0 Introduction

Even highly automated vehicles (AV) are likely to encounter problems in their programming which may require intervention, in the form of a human operator assisting from a remote location. These scenarios may be more frequent than one might first think as a host of 'edge cases' related to AV perception issues, such as poor traffic light perception, adverse weather, debris in the road or unexpected behaviour of other road users have already been documented by automated driving companies as causes of AV disengagement from the driving process (California DMV 2019). Humans are capable of interpreting and reacting to novel scenarios which will not necessarily be anticipated within the AV’s programming. The functional scope that an RO may be asked to perform could be 1) remote assistance, merely advising passengers of delays or opening and closing doors, 2) remote supervision where they may provide instruction to the AV in the event of navigational failure or ambiguous perceptual information or 3) remote driving, where they will be required to take active control of all the driving, steering and braking functions of the AV (Mutzenich, Durant, Helman, & Dalton, 2021). Although a Remote Operator (RO) will rarely be expected to take dynamic control of the AV's driving functions, they are likely to regularly be called upon to carry out remote supervision tasks such as evaluating the environment using the AV’s cameras and sensors and providing human direction whether it is safe to proceed or advising what alternative navigation is required {Formatting Citation}. ROs of AVs will be unable to provide remote supervision until they gain necessary situation awareness. Situation awareness (SA) refers to perceiving and understanding the environment and how it may change in the near future, in the case of an RO from a "far-off" location (Mutzenich et al., 2021b, p. 7).

Currently, without known exception, AV companies that offer RO assistance (for example, Phantom Auto in the US, Imperium Drive in the UK and German based company Vay amongst others worldwide) provide SA information to ROs via monitor feed from sensors such as cameras and lidar to deliver information about the environment from around (and even above) the remote vehicle. This information is relayed from video feed to as many as 6 screens in front of the remote operator, enabling them to control and manoeuvre the machine from a remote location (Linkov & Vanžura, 2021). Monitor-based video presentation is limited to the view that the fixed cameras on the AV afford, which may require time-consuming manual control to ‘look around’ the remote scene by accessing different camera feeds to gain spatial awareness (Jankowski & Grabowski, 2015).

An alternative to traditional approaches of screen-based video presentation to facilitate remote operation, could be to represent the video feed from the vehicle's location via a head-mounted display (HMD) where field of view (FOV) changes are based on head movement. This creates the illusion that the user is viewing the remote environment from an egocentric perspective, and that they only need to turn their head to activate the same movement in the AV cameras (Mutzenich et al., 2021b).

Two critical questions for the industry concern: 1) What information ROs will need to build and maintain SA quickly and safely, and 2) in what format this information should be delivered. The focus of this paper is the second question, where we consider two modes of presentation to carry out remote supervision of an AV; either a conventional computer monitor view, or a head mounted display (HMD) both showing videos recorded in 360°.

1.1 Head mounted display presentation versus screen-based monitor presentation in remote operation

Although there have been few studies investigating the use of HMD in remote driving scenarios, other industries have long been making use of its benefits, for example in situations where real conditions are dangerous for humans, such as in mining (Grabowski & Jankowski, 2015), for enhancing gaming experience in 360° first person shooter games (Monteiro, Liang, Wang, Chen, & Baghaei, 2020) or in robot teleoperation (Zhang, Minakata, & Hansen, 2019). HMD showing 360° videos has also been used in driving training contexts to assess drivers' hazard perception skills as a viable alternative to current hazard perception tests on monitor screens which artificially concentrate the viewer's focus to impending hazards directly ahead (Crundall et al., 2021).

Whether HMD could be a feasible mode of presentation in remote driving situations compared to conventional screen-based presentation has been evaluated using virtual driving simulations with limited driving control. Weidner, Hoesch, Poeschl, & Broll (2017) investigated how different display types on either a monitor (2D in 360° or Stereoscopic 3D) or an HMD (360°) influenced driving performance of a Lane Change Task. They found that HMD did not result in differences in driving performance but significantly increased simulator sickness. Presenting video information in HMD can result in cybersickness symptoms such as nausea, fatigue and headaches as a result of visual motion being perceived through electronic screens even though our physical body remains inert (Chang, Kim, & Yoo, 2020). However, many situations which require RO assistance will involve interaction with an AV which has come to a safe stop meaning in these use cases that cyber sickness may not be a barrier to adopting HMD presentation as the vehicle is stationary.

Research has also considered whether HMD presentation is superior to viewing remote videos on conventional monitor set-ups using real world remote driving tasks. Georg, Feiler, Diermeyer, & Lienkamp (2018) investigated the effect of HMD vs. conventional monitor presentation on telepresence and workload when remotely operating a test vehicle in the real world. Although test subjects reported higher satisfaction with the HMD, they still had several collisions in narrow city scenarios in HMD conditions which suggest that actual driving performance is not improved by HMD presentation. However, the relative superiority of HMD presentation may depend upon which RO use case is required, as not all RO tasks demand direct driving control, for example advising the AV of the on-going direction in their route in the event of navigational failure.

Research has explored the effect on supervisory based tasks when virtual driving environments are presented in HMD compared to standard monitor view. Georg et al. (2020) measured the impact on situation awareness and decision making, presenting information in high or low streaming quality in either monitor presentation or via a virtual environment presented in HMD in 360° so the operator could look around as if they were inside the vehicle. The Situation Awareness Global Assessment Technique (SAGAT), a freeze and probe method, was used to test participants' perception of road users, comprehension of traffic signs and whether it was safe to continue the journey at the end of the scene. Results showed significant differences in SA between the display formats, with participants' decision whether to continue the journey being more accurate if presented in higher quality streaming but reported no significant findings between the different display modes. The HMD interface had higher scores of subjective satisfaction, even though the conventional monitor received higher confidence ratings when participants were asked to envisage using it as an RO driving in real road traffic. Bout, Brenden, Klingeagrd, Habibovic, & Böckle (2017) also employed simulations to compare 360° videos presented in either HMD or screen-based display of pre-recorded traffic scenarios, for example depicting an AV which has stopped due to an obstruction in the road. They found that HMD presentation facilitated easier interpretation of the traffic situation although feedback indicated that participant's preference between an HMD or computer display was dependent on scenario. These findings suggest that HMD presentation has some potential to have a positive effect on gaining the SA which ROs require to supervise decisions about the next course of action for an AV to take but conventional monitor view may still be favoured dependent upon context and use case.

1.2 Presence in remote operation

Presence (sometimes referred to as telepresence in the context of remote operation) has been described as a subjective perception of being somewhere other than your physical location (Georg et al., 2020). Studies have found positive relationships between presence and task performance in virtual environments with spatial presence and a sense of 'realness' as the strongest components that determine a "sense of being there" (Schubert, Friedmann, & Regenbrecht, 2001, p. 3). Possessing a sense of presence for an RO, would be the sensation of being at the AV's environmental locale rather than at the operator's control station (Witmer & Singer, 1998). Information presented via video feed on screen-based monitors may result in detachment or a decreased analysis of risk as a result of a diminished sense of presence in the remote scene. Consideration of the effect of variations of FOV on operator performance and embodiment through manipulation of screen based presentation or HMD have found that HMD techniques largely have a positive effect on operator efficiency and sense of presence or telepresence in the virtual world (Grabowski, Jankowski, &

Wodzyński, 2021; Monteiro et al., 2020; Zhang et al., 2019). If an RO has to adopt a supervisory role over an AV, feeling present in the remote scene will be critical to fully engage with the task (Simpson, Bolia, & Draper, 2013).

It seems likely that HMD presentation has the potential to improve an RO's sense of presence in the remote scene. Research that has compared HMD interfaces with screen-based presentation have found participants give higher ratings of presence for HMD, even though conventional monitor presentation received higher confidence ratings when participants were asked to envisage using it as an RO driving in real road traffic (Georg et al., 2020). However, Georg, Feiler, Diermeyer, & Lienkamp (2018) compared HMD with a monitor-only setup when actually driving an teleoperated vehicle in the real world and found no significant difference in reported sense of presence between the two setups. Participants reported that their perception of reality was higher in the conventional monitor view than the HMD, which researchers attributed to the high latency rates within the HMD interface undermining the confidence that users could place in the visual feed from the car. This suggests that methods which will provide ROs with a realistic sense of presence are worthwhile considerations for remote supervision, but only if they are matched by improvements in technological latency for remote visual feeds.

1.3 Eye tracking measures as an indicator of remote operator SA

Inadequate SA is frequently implicated in highway crashes and failed to look/distraction is the most common citation in driving insurance documentation (Department for Transport, 2014). Existing SA measurements using "freeze and probe" quantitative techniques, such as the Situation Awareness Global Assessment Technique (SAGAT) have been criticised as unsuitable in dynamic, uncontrolled environments and measurement of SA in remote contexts may be more challenging still (Endsley, 1988). Furthermore, SAGAT relies on what the operator has been able to remember and then explicitly recall rather than a true measure of ongoing awareness of the environment (de Winter, Eisma, Cabrall, Hancock, & Stanton, 2019). Eye movements have been raised as a potential real-world measure of SA, in line with the 'eye-mind hypothesis' (that where a person is looking is an indicator of what they are attending to), which go beyond perception of hazards and can show how operators allocate attention, giving insight into their SA (van de Merwe, van Dijk, & Zon, 2012).

Fixations measure periods of time where the eyes are still whilst they take in information. In the context of presentation of information from the location of an AV to an RO, more frequent fixations on areas of interest (AOIs) in a visual scene may predict good *perception* for those areas of the driving environment yet, alternatively, wider spatial distributions of fixations may signal better *understanding* of the scene due to superior scanning behaviours (Shahar, Alberti, Clarke, & Crundall, 2010). This suggests that measuring eye movements while operators make supervisory decisions based on the remote context could provide information about the visual search strategies that are used to gain awareness of remote scenes and can provide insight into the visual cues that people draw upon to make SA decisions in remote contexts.

There is debate about what eye movements can tell us about SA; although we can see where someone is looking, we cannot know if they are attending to the visual information in that location. Tasks that require participants to passively view driving videos may result in visual behaviour that conflicts with eye movements found when actively driving, for example visual search for hazards in real life may be lower due to increased attentional demands (Mackenzie & Harris, 2015). Furthermore, driving in real life requires us to pick up locomotor cues at high speeds using active gaze control to perceive hazards ahead of us and adjust steering accordingly, this will not be necessary for an RO carrying out a supervisory task as the AV will be stationary (Wilkie, Kountouriotis, Merat, & Wann, 2010). Consequently, eye movements during visual search for SA information in an RO supervisory task may not be comparable to eye movements during remote driving tasks.

In a remote supervision task, ROs with good SA would be expected to fixate on relevant information in the remote scene that contribute to the understanding and projection of current and future traffic situations. However, depending on how the information is presented to ROs there may be differences in distributions of fixations. Studies of visual perception have found that the best placement to extract visual information from a computer screen is the centre display as observers have a systematic bias to look at this location first (Bindemann, 2010). Eye tracking research, when studying web page design, has observed that people are inclined to prefer elements on the left hand side of the page when searching for goal-directed information whereas the top right received almost no visual attention at all (Buscher, Cutrell, Morris, & Way, 2009). This suggests that if computer monitors are used to present the video feed from an AV, ROs may distribute fixations differently than if the same information was presented in an HMD.

1.4 Study overview

The aim of the current study is to investigate remote supervision using a choice decision task to assess the effect of two formats of presentation of 360° driving video, either head mounted display (HMD-360 condition) or screen-based monitor display (SB condition), when carrying out remote supervision of automated vehicles. Participants viewed 60 videos, filmed from the starting perspective of a stationary car in 360°, either presented in a head mounted display (HMD-360) where field of view (FOV) could be changed by moving their head or via screen-based (SB) presentation where FOV was controlled by using the mouse to move the image around on-screen. The experimental task required participants to decide the most appropriate on-going direction of travel, out of one of four options on a keypad, using SA information from the scene. They were not expected to conduct the manouvres. The dependent variable was the time taken in seconds to press the corresponding button on a keypad indicating their decision.

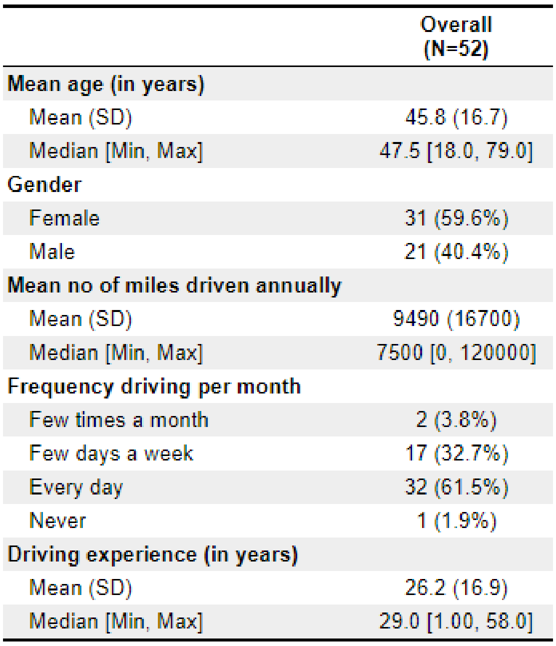
We tested two behavioural hypotheses; 1) that presentation format (SB/HMD-360) will affect decision time and 2) that presentation format will affect accuracy of decision. We also conducted exploratory analyses on whether the type of video scenario affected decision time and accuracy and whether the type of decision (continue, left, right or reverse) impacted on decision time and accuracy. We further tested an eye tracking hypothesis that the distribution of fixations on AOIs in the video scenes will be different across conditions. Finally, we predicted that there would be differences between the two conditions in terms of self-reported cyber sickness symptoms, sense of presence and workload.

2.0 Method

2.1 Participants

Participants were recruited either via the paid participant pool at Royal Holloway university Psychology department or in response to a recruitment campaign by Transport Research Laboratory (TRL) sent out to 500 local residents. Recommendations for power analysis for a mixed effects model indicated that a sample size of at least 52 participants was needed to result in power of 0.8 (α-level = 0.05) to report medium effect sizes (Brysbaert, 2019; Brysbaert & Stevens, 2018)

Table 1 Demographic data from main study n=52



	Overall (N=52)
Mean age (in years)	
Mean (SD)	45.8 (16.7)
Median [Min, Max]	47.5 [18.0, 79.0]
Gender	
Female	31 (59.6%)
Male	21 (40.4%)
Mean no of miles driven annually	
Mean (SD)	9490 (16700)
Median [Min, Max]	7500 [0, 120000]
Frequency driving per month	
Few times a month	2 (3.8%)
Few days a week	17 (32.7%)
Every day	32 (61.5%)
Never	1 (1.9%)
Driving experience (in years)	
Mean (SD)	26.2 (16.9)
Median [Min, Max]	29.0 [1.00, 58.0]

Participants self-selected their involvement in the study by responding to the e-mail and booking an appointment slot online. Testing took place in one of three locations: either in the VR lab at Royal Holloway, Egham, the offices of TRL in Crowthorne, Surrey or in a temporary laboratory in Godalming, Surrey dependent on participant location. The lab set-up for each location was the same and experimental materials and hardware were kept constant for each location. Participants gave informed consent before participating, were paid £10 per hour for taking part in the study and those who attended TRL were paid £5 for travel costs. All procedures were reviewed and approved by the Royal Holloway Research Ethics Committee. The study took between 30-45 minutes to complete.

All participants were UK residents with English as their first language and possessed a full UK driving license. Table 1 gives demographic details relating to the frequency of driving, number of years they had held a license and their gender. The average age of participants was 45 years and more female participants responded to the recruitment campaign. Over 90% of the participants drove either every day or a few days a week with an average of 26 years driving experience.

2.2 Materials

2.2.1 Creation of driving videos

To create the driving videos[1], the lead researcher drove a 2011 BMW 120 diesel car (manual right-hand drive) with a Ricoh Theta Z1 fixed to the car on the roof at 27cm lateral and 20cm longitude to record 360-degree raw video at 4K resolution. Recording was controlled inside the vehicle using a mobile phone link to the camera. All videos were filmed within the GU7 postcode in Surrey, England and on campus at Royal Holloway University, Egham during a two-day period in October 2021, selected for dry weather conditions.

No videos were filmed on motorways or dual carriageways due to restrictions on parking and to ensure the safety of the researcher and other road users. Video 'scenes' were contrived using existing traffic events such as roadworks and construction works on the Royal Holloway campus or existing signage (such as no right turn signs) on public roads such as cones blocking off entryways. No members of the public were identifiable in the films due to the resolution of the video presentation. A limited number of videos which featured people using hand gestures to direct traffic were produced with the assistance of private construction workers working locally at the time of recording who gave full permission to be filmed.

The lead researcher identified potential filming opportunities, parked the vehicle, and started the roof-top camera from inside the car, filming for 30 seconds with hazards lights on, before stopping the film and moving off. All videos were presented starting from the point of view of the bonnet of

---

[1] https://github.com/Anthrometric/remote_supervisor_stimuli_driving_videos.git

the car to show the direction the car is travelling in. The experimental task required participants to make an active decision as to the direction the vehicle would need to take next (left, right, continue forward or reverse), using the information from the video. Ninety films were created to present the experimental stimuli response choices of turning turn left or right, continuing forward or requiring reverse manouvres.

Videos were classified into seven groups according to the type of edge cases that they involved (all of which are likely to be presented to ROs in real world situations). The different types of edge case were spread across the different decision choices (left, right, forward or reverse) as far as was pratically possible given the constraints involved in capturing appropriate video footage. Table 2 shows the different types of scenarios and how many videos were presented in that group.

Table 2 Scenario classifications and number of videos in each group

Classification	Diverted route	Construction works	Highway code comprehension	Closed gates	Object in road	Dead end	Signs/signals
	e.g. cones blocking one route option	e.g. traffic lights or signs indicating new road rules	e.g. passing parked cars on the side of the road by moving into the opposite lane, waiting at railway crossings etc.	e.g. route bocked by closed and fastened gates	e.g. scooter, plastic bag	e.g. cul de sac, fence	e.g. on road signs or road way signs and signals such as no entry, right turn only etc.
No of videos in set	9	10	12	6	4	9	10

2.2.2 Post-recording processing

To play the videos as 360 videos on a 2D screen, we injected spherical metadata information using a Spatial Media Metadata Injector script[2]. The metadata injection is automatically done in Tobii Pro Lab for the VR presented videos. Original resolution directly from the Theta camera was 3840 x 1920, but as a result of repeated crashes of the Tobii Pro Lab software in the VR condition during piloting, we used a downsampled resolution of 2048 x 1024 at 30Hz using Handbrake (version 1.4.2) for the VR headset and adopted the same resolution (2048X1024) for the desktop-based 2D settings to ensure that the that the spatial information was the same between the two conditions. All videos were cut to 29 seconds in length using using Camtesia©[3] with a 5 second fixation cross added to the start. Video audio was muted due to the variations in salient audio cues between some of the videos which may have assisted in the decision process.

---

[2] https://github.com/google/spatial-media/releases/tag/v2.0

[3] www.techsmith.com, version 2018

2.2.3 Creation of playlists

The four-person research team viewed all 90 videos independently and rejected 22 videos in which the correct judgement was contentious. Eight videos, two examples of each possible answer (left, right, forward, reverse) were selected as practice videos. The remaining 60 videos were divided into two playlists with similar ratios of correct answers (see Table 3) and balanced by video classification, for example videos which contained diverted routes, construction, closed gates etc. were represented equally in each playlist to ensure that one playlist did not have more examples of certain scenarios than others, as some may prove easier to make decisions about than others. See Appendix 1 for a full list of video names, description of the scenarios and correct answers.

Table 3 Details of correct answers for each playlist and the number of videos that corresponded to each answer in Playlists A and B.

Correct answer	Total number of videos	Playlist A	Playlist B
Continue straight ahead	18	9	9
Left	7	3	4
Right	16	7	9
Reverse	19	11	8

2.2.4 Hardware and lab set-up

The experimental set-up and hardware were identical in each of the three testing locations. The HMD-360 condition utilised the HTC Vive VR-HMD with embedded Tobii Pro Lab eye-tracker to present the videos in either Playlist A or B in random order to participants at 90Hz display refresh rate and a screen resolution of 1080x1200 pixels per-eye (120 Hz gaze data output frequency). Their FoV was around 90° horizontally and vertically (although it slightly changes between participants dependent on head size). They were seated next to a table in a fixed chair to limit virtual reality sickness with the keypad situated on the table to their left-hand side within comfortable reach.

In the SB condition, participants were seated at a Dell (S2719H) 27-inch monitor (display width about 23.62" (60 cm) and display height about 13.39" (34 cm) with a resolution of 1920x1080 pixels at 60HZ display refresh rate and FoV around 60° horizontal and 34° vertical. They had a keyboard in front of them, a mouse on their right-hand side and a keypad on the left-hand side. Eye tracking was calibrated using the Tobii Pro Nano screen-based eye-tracker[4] which records gaze at 60Hz. The experimenter sat on the other side of the screen using a workstation that mirrored the monitor so they could set up the eye tracker and start the relevant playlist in the SB condition. The experimenter used VLC Media player, an open-source media player software, to present either Playlist A or B in random order to the participant and manually recorded the order of presentation

---

[4] https://www.tobiipro.com/product-listing/nano/

for each participant. Tobii Pro Lab screen recording was activated to overlay participant gaze data onto each video so the replay function could be used for checking validity of coordinates.

The keypad was used for response collection in both conditions with tactile indicators attached to the buttons to reduce the likelihood of participants pressing incorrect buttons, particularly in the VR condition when the headset impeded a visual verification of the button pressed. Coarse Velcro was cut into arrow shapes to represent directions for the four choices in response to the videos (continue forward [8], go left [4], go right [6] and reverse [2]). The middle button [5] was a square-shaped soft Velcro to distinguish it from the other response options and was only pressed in the SB condition to move to the next video. See Figure 1 for visual details of the experimental setup.

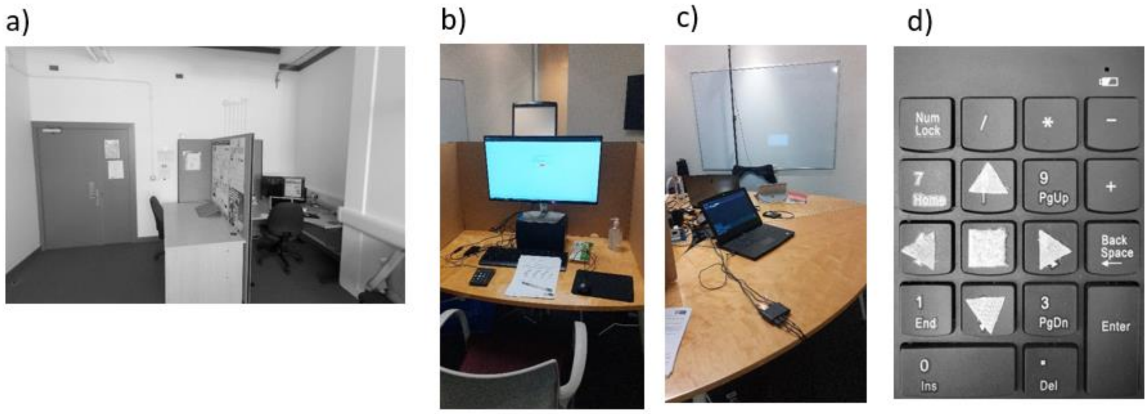


Figure 1 Experimental set up of lab showing a) the experimenter [left] and participant position [right], b) the (left-hand) keypad and (right-hand) mouse position for the SB condition with 17" Dell monitor , c) the position of the keypad on the table (viewed from the perspective of the participant's seated position) for the VR condition and d) the keypad used in both conditions with tactile indicators for responses (continue forward [8], go left [4], go right [6] and reverse [2]). The middle soft button was pressed in the SB condition to move to the next video.

2.3 Design

The study employed a within subjects design with one fixed factor (mode of presentation: SB and HMD-360) and two random factors (participant and video) analysed by a linear mixed effects model (LMM). We presented 60 videos starting from the forward-facing perspective of a static car, half of which were viewed on a monitor in 360° (SB condition), with the other half viewed in a VR headset (HMD-360 condition). Participants in the SB condition could look around the video environment by clicking on the screen with the mouse and moving the video around 360°. In the HMD-360 condition, they could look around the 360° video by moving their head.

All 52 participants saw all 60 videos once in either of the two conditions. Presentation of playlists A and B in each experimental condition (HMD-360/SB) was counterbalanced across participants, as was the order of presentation of the conditions (shown in Table 4). The presentation of videos within each playlist was randomised for each participant.

2.4 Procedure

Participants arrived, received a short oral briefing and filled out a demographic questionnaire online which was hosted at Gorilla.sc. They were told the order in which they would be completing the experimental conditions and asked to move to the area of the room that housed the first condition. In the HMD-360 condition, they were handed the headset and given instructions how to clean it and place it on their head, making any necessary adjustments until it was comfortable (COVID-19 restrictions limited us from doing this step ourselves). In the SB condition they were seated at the workstation in front of the monitor. In both experimental conditions, the position of their eyes was calibrated, and they were shown instructions outlining the task on-screen.

In each condition, participants watched 30 videos presented from the viewpoint of a static car and were told that they were viewing the video starting from the perspective of an AV that had stopped facing in the direction it would ideally go in enroute to its destination. The task was to assess each video scene and decide the most appropriate on-going direction of travel, indicating their decision by pressing the corresponding button on the keypad using their non-dominant hand. All the problems depicted in the videos could be answered with one of four options on the keypad (continue forward [8], go left [4], go right [6] and reverse [2]).

A practice trial was carried out before the experimental trials, using the same 8 videos in both conditions, showing examples of when each decision option would be the correct answer. This was to familiarise the participant with both modes of presentation, to practise using the keypad to indicate their decision, to practise moving the 360° images around with the mouse in the SB condition, and to ask the experimenter any questions.

Table 4 Presentation order of playlist by condition

Presentation order of playlist and condition	
Playlist A - HMD-360 condition	Playlist B - SB condition
Playlist A - SB condition	Playlist B - HMD-360 condition
Playlist B - HMD-360 condition	Playlist A - SB condition
Playlist B - SB condition	Playlist A - HMD-360 condition

Participants were advised that they could make their decision as soon as they knew what to do, even if it was not possible for the action to be carried out immediately as the AV would carry out the instruction once it was safe to do so. Full participant instructions can be found in Table 5.

Table 5 Participant instructions for the experimental task

Participant instructions
*The AV may have stopped because something is in its way - if you are able to go around it, you can **go forward** even if this involves briefly going onto the other side of the road such as when overtaking parked vehicles.* *If the problem will not change in the next few seconds, for example there is an obstacle permanently blocking the way or road traffic information instructing a change of direction, you may have to **go/turn left** or **go/turn right** or **reverse**.* *Reverse is always the last option when there is no possible way ahead, for example the road is blocked, with no possible right or left turns or is unsuitable for a car to drive on. **Don't worry about the direction the AV will take once they have finished the reversing manouvres** necessary to get them out of the current situation, you can just indicate **reverse** as the decision.*

At the end of each condition, participants completed the NASA TLX-R Workload measures (Hart, 2006) indicating answers on a sliding scale where responses were quantified from 'very low' (0) to 'very high' (100). To measure their sense of presence in that condition, four questions were adapted from Schubert, Friedmann, & Regenbrecht's (2001) presence scale. One question, " *How real did the virtual world seem to you*?" was answered on a sliding scale from 'completely real' (0) to 'not real at all' (100). Three questions asked participants to indicate their level of agreement with the following statements, "*In the virtual scenes I had a sense of 'being there'*", "*I was completely captivated by the virtual scenes*", and " *I felt present in the virtual scenes"* and each response was measured on a Likert scale from "strongly agree" (0) to "strongly disagree" (5). Questions about any symptoms of cyber sickness they were suffering from were adapted from the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, 1993) with each item rated on a scale of none [0], slight [1], moderate [2], severe [3]. Figure 2 shows the full procedure of the study.

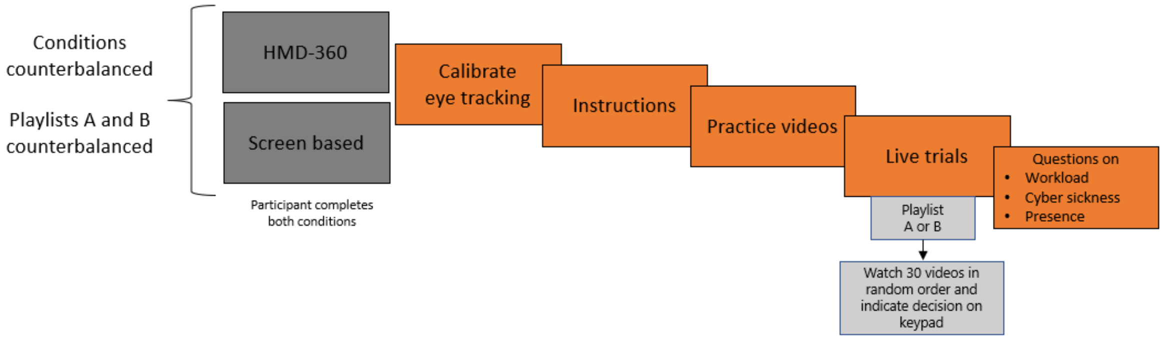


Figure 2 Procedure of the experiment n.b. Participant completes all tasks depicted in orange boxes in both conditions and answers questions at the end of each condition

3.0 Behavioural results

3.1. Measurement of decision time

For Analysis 1, which tested the prediction that presentation format (SB/HMD-360) will affect decision time (DT), in the SB condition, the dependent variable of DT was calculated as the time between the number 5 keypress (to start the new video) and the next keypress (2,4,6,8), minus 5 seconds (for the fixation cross). In the HMD-360 condition the video started automatically so DT was defined as the time from the start of each stimulus to the keypress minus 5 seconds. For Analysis 2, which tested the prediction that presentation format will affect accuracy of decision, the correct answer for each video was scored as 1 and 0 if incorrect.

3.2 Data cleaning decisions

Each presentation of a video stimulus was recorded as a trial. Participants had a total of 30 trials in each condition and one keypress was recorded as a response to each trial. The time between the start of the video and the keypress decision was recorded as decision time (DT) with 5 seconds subtracted for the duration of the fixation cross at the start of each video. The maximum DT length was 29 seconds as all videos timed out after that duration. Manual recording of order of videos for each participant was added to datasets in the SB condition. This information was automatically recorded by the Tobii Pro Lab software in the HMD-360 condition.

Any trials which were recorded as NA were excluded. Any keypress that was not a decision option (defined as a 2, 4, 6 or 8 keypress) was defined as an incorrect response and the DT for this video was recorded as NA. If the participant held the keypress down for too long, it resulted in multiple keypresses being recorded and the next video being skipped so DT for that trial were also recorded as NA. Any videos which were not presented to participants due to software or errors were also recorded as NA. The original sample consisted of 52 participants, but data for five participants proved invalid due to software/computer faults or data storage errors.

3.3 Justification of analysis

We used a linear mixed-effects model (LMM) rather than repeated measures analysis of variance (ANOVA) as the effect of participant and the effect of the stimuli (in this case the videos shown to participants) can be set as random effects which partitions the variance associated with these differences explicitly. For example, some subjects may be faster than others at the task or may get faster, or some videos may be perceived as more difficult than others. Secondly, in this study, some participants pressed the button too quickly and missed videos or pressed a button on the keypad that was not one of the response options and so this data was recorded as NA with no decision time data giving an overall data loss of 6 observations (HMD-360) and 13 (SB) out of a potential 1410 observations in each condition. LMM allows for missing data whereas traditional repeated measures ANOVAs do not, which gives a clear advantage of LMM analysis over ANOVA for this study.

3.4 Analysis of results

Initial examination of descriptive statistics (see Figure 3) suggest that the time take to make a decision was slower in the SB condition. On average, participants took around 5.81 seconds to make decisions in the SB condition compared to 4.65 seconds in the VR-360 condition.

Participants were more likely to make accurate decisions in the VR condition, although decisions made in the SB condition were still highly accurate.

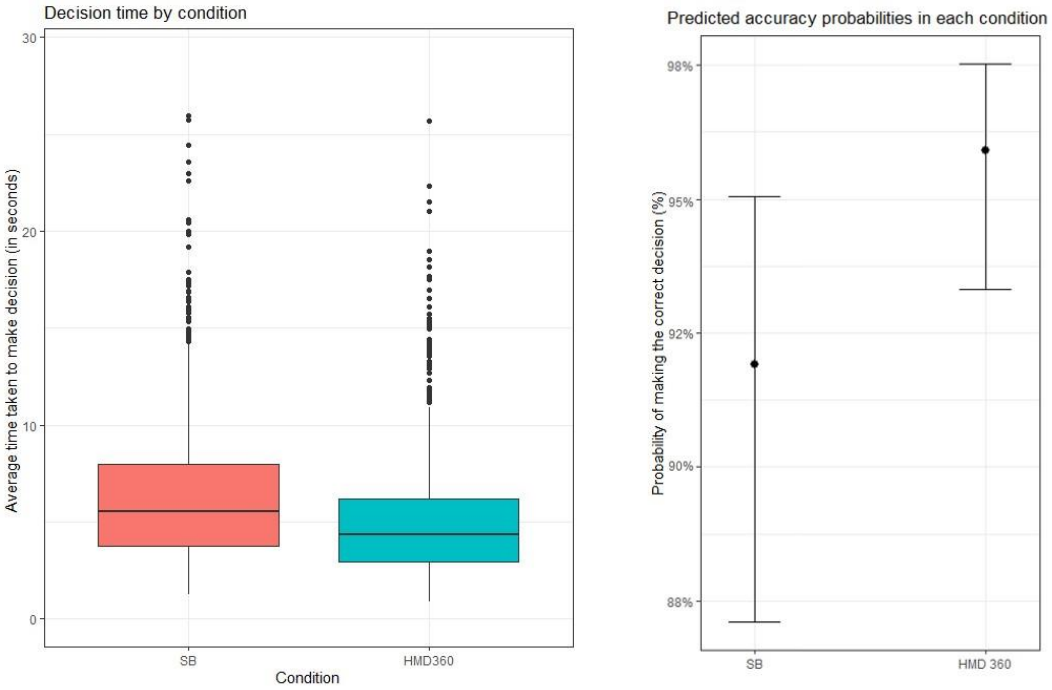


Figure 3 Descriptive statistics to show (left) decision time by condition and (right) the predicted accuracy probability for each condition

3.4.1 Analysis #1 - Presentation format (SB/ HMD-360) will affect decision time

To test the hypothesis that presentation format (SB/ HMD-360) will affect decision time, we ran a linear mixed effects model in R and used lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) to calculate significance which generates p values using Satterthwaite's method. Condition (SB/HMD-360) and Trial (1-30) were entered as a fixed effects and we included Subject and Video as random effects. The decision time (DT) to press the decision button was the dependent variable.

We adopted a maximal model strategy in line with recommendations by Barr, Levy, Scheepers, & Tily (2013) which included single by-participant random slopes for the effect of condition to incorporate variations in how different participants may respond to the SB/HMD-360 manipulation. Finally, we included an interaction term between condition and trial to test whether there was an interaction between when they saw the video over time[5]. Visual inspection of plots suggested that the data was skewed leaving a poor fit of the model, so to satisfy model assumptions, DTs were log transformed and used for all following analyses. Bootstrapping was used to estimate confidence intervals based on 500 simulations.

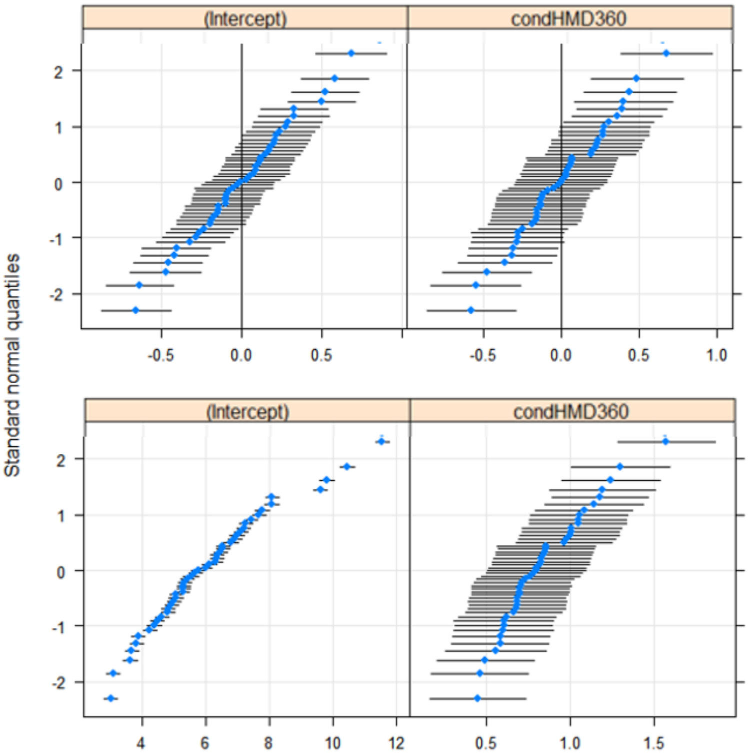


Figure 4 Analysis #1 Random effects coefficients of the model for subject plotted against normal quantiles show how the by-participant intercepts vary from overall average when the intercept is centred around zero (top panels).

To show the difference the HMD-360 condition makes for each participant (bottom right panel) we have added the fixed effect term (1.4) from the model. The bottom left plot shows the intercept for each participant. This allows us to see how averages vary for each participant. X-axis shows transformed log seconds.

Results showed there was a significant main effect of Condition (log DT β = -.22, t = -4.031, p <0.001,95% CI []) and a significant main effect of Trial (log DT β = -.004, t = -2.241, p <0.03,95% CI [-0.330, -0.160]) with no interaction (t<1). Log DTs were exponentiated to present changes in multiplicative terms. On average, decisions were made in the HMD-360 condition in 80% of the time required in the SB condition. The average difference in log DT over 30 trials was around 96%. Figure 4 shows the difference the HMD-360 condition makes in the log DT for each participant.

To estimate the relevance of the predictor, we compared our model to a simpler one without the fixed effect of interest with only an intercept term (and without an effect of Condition or Trial). We

---

[5] *lmer(log DT ~ cond*trial +(1+cond*trial|subj) + (1+ cond*trial|video), data = combined_results)*

used the anova() function to check whether inclusion of the mode of presentation (HMD-360 /SB) fixed factor and trial (1-30) significantly improved model fit. Model comparison through aka likelihood ratio test demonstrated that the full model with random slopes and intercepts significantly improved model fit (X^2 (20) = 432.76, p=0.001) meaning there is a significant main effect of Condition and a significant main effect of Trial.

3.4.2 Analysis #2 - Presentation format (SB/HMD-360) will affect accuracy of decision

It is important that remote operators are not only quick to make decisions but that those decisions are based on accurate situation awareness. We specified a single binomial response for each trial, reflecting whether the participant correctly chose the right decision option (1) or not (0). Timed out responses for trials were not included. We fitted a model[6] with random intercepts for participant ID and Item (videos) and specified a generalised linear mixed effects regression using the glmer() function with instruction to treat the data distribution as binomial with random slopes for each participant, and for each item. We included single by-participant random slopes for the effect of condition to incorporate variations in how different participants may respond to the SB/HMD-360 manipulation. The dependent variable was the accuracy of decision for each video.

Results showed a significant effect of Condition (β= 2.695, SE=0.27, z(2727)= 2.70, p<0.01, CI 95%[1.18, 3.78]). We exponentiated the model coefficients (0.72) which revealed that the odds ratio of making a correct decision in HMD-360 was over twice that in SB (2.06). Finally, to investigate the difference the HMD-360 condition is making in the log of the odds of the outcome we ran the ploggis() function in R which showed the likelihood (probability) that someone in HMD-360 condition made an accurate decision was 96%, although the SB condition was still highly accurate at 92% (see Figure 3).

3.5 Exploratory analyses

3.5.1 Exploratory analysis #1 - Effect of video type on decision time

Some of the 60 videos may have been easier to make decisions about than others as they covered a broad range of edge case scenarios. To reduce parameters to fewer levels and to lower the variability of the model, we fit a linear mixed effects model which categorised seven types of scenario represented by the video (see Figure 2 in Section 2.2.1). The dependent variable was the time to press the decision button (log_DT). Type 1 (Construction videos) were chosen as the baseline

---

[6] *glmer(correct ~ cond + (1+cond|subj) + (1+cond|video),data = combined_results,family = binomial)*

as the workmen depicted in the videos signalled the direction to take via hand gestures so this video type represented the most undemanding decisions so decision times may have been faster.

We specified a maximal model which included Condition (HMD-360/SB) and Video type (Type 1 [Construction], Type 2 [Dead ends], Type 3 [Diverted route], Type 4 [Gates], Type 5 [Highway code], Type 6 [Object in the road] and Type 7 [Signage]) as fixed effects as well as an interaction between Condition and Video type. We added Subject as a random effect, with by-participant random slopes[7].

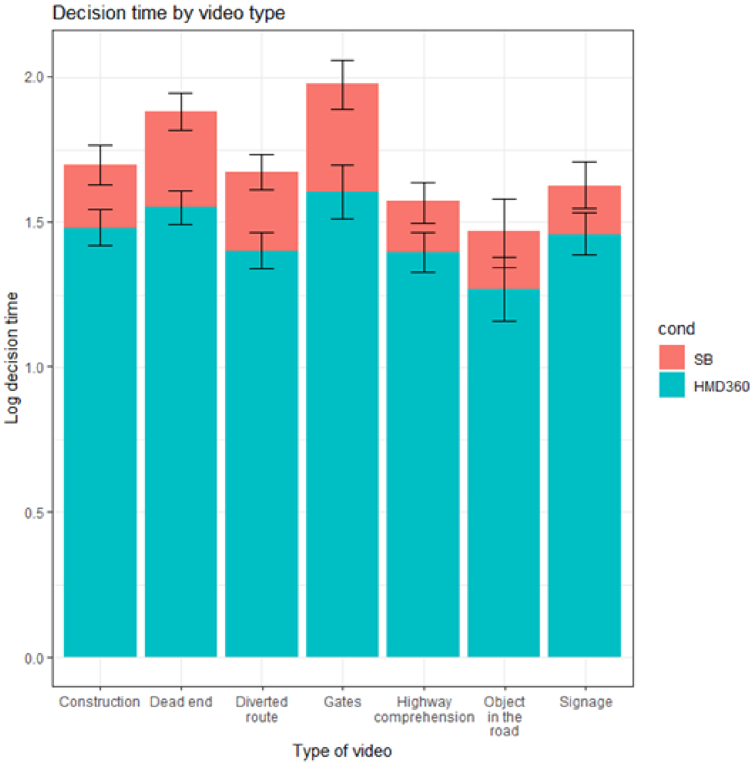


Figure 5 Decision time by video type in each condition

We found a significant main effect of Condition (log DT β=-0.25, SE = 0.05242, t=-4.699, p 0.001). Results showed a significant main effect of Type 2 [Dead ends] (log DT β=0.15, SE = 0.04877, t=3.156, p= 0.01), Type 4 [Gates] (log DT β=0.22830, SE = 0.06, t = 4.07, p = 0.001), Type 5 [Highway code comprehension](log DT β=-0.13, SE 0.05, t= -2.954, p =0.01), Type 6 [Object in the road] (log DT β= -0.25, SE 0.06, t = -3.84, p = 0.001) and Type 7 [Signage] (log DT β= -0.10, SE = 0.04, t= -2.20, p= 0.05) with no significant interaction effects (t <1 for all Video types and Condition).

Log DTs were exponentiated to present changes in multiplicative terms. As previously found in Analysis 1, on average, decisions were made in the HMD-360 condition in 80% of the time required in the SB condition. For the fixed effect of video type, in comparison to the baseline of the time taken to make decisions for Type 1 [Construction] videos, participants on average were slower at making decisions for Type 2 [Dead end] videos (about 66% increase in performance time) and for Type 4 [Gates] videos (about 26% increase in performance time), which suggests that participants found the types of scenarios presented by these videos more challenging. Conversely, the time

---

[7] *lmer(log DT ~ cond*Type + (1+cond*Type|subj),data = combined_results*

taken to make decisions in response to Type 5 [Highway], Type 6 [Object in the road] and Type 7 [Signage] videos were all between 78-91% of the time taken for the baseline videos.

3.5.2 Exploratory analysis #2 - Effect of video type on accuracy of decision

Another area of interest was whether the type of edge case scenarios presented to participants through the videos used in our study would influence accurate decisions. We fitted a similar model to Analysis 2, using a generalised linear mixed effects regression model[8] with instruction to treat the data as binomial, with Condition and Video type as fixed effects and we added by-participant random intercepts and slopes. The dependent variable was the accuracy of the decision for each video. We kept Type 1 (Construction) as the reference condition.

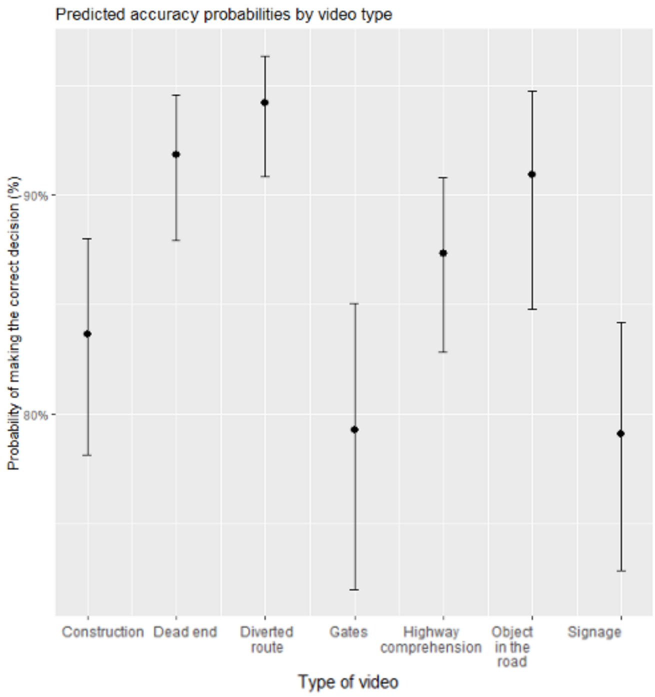


Figure 6 Probability of accurate decision by video type

We found no significant effect of Condition ($t = 1.91$, $p = .1$), with significant main effects of Type 2 [Dead end] ($\beta=-0.75$, SE = 0.24, $z = 3.12$ $p = .02$) and Type 3 [Diverted route] ($\beta=-1.20$, SE = 0.27, $z = 4.47$, $p = .001$). Exponentiated model coefficients showed that the odds ratio of making a correct decision in comparison to Type 1 [Construction] videos was over twice (2.12) for Type 2 [Dead end] videos and over three times (3.30) for Type 3 [Diverted route] videos. The likelihood (probability) of making a correct decision was 93% for Type 2 videos and 95% for Type 3 videos compared to 86% for the Type 1 videos showing that participants had better SA performance when videos showed dead ends or diverted routes.

---

[8] *glmer(correct ~ cond + Type + (1|subj), data = combined_results, family = binomial)*

3.5.3 Exploratory analysis #3 - Effect of decision type on decision time

In this study, some decisions in response to the task, such as 'Continue' may be made more rapidly than others such as 'Reverse', which may require the other 3 options to be exhausted before coming to the keypress choice. To investigate whether there was a main effect of decision 'type' on log DT, we fit a linear mixed effects model with Condition and Decision type (Continue, Left, Reverse, or Right) as fixed effects with an interaction term, adding Subject and Video as random factors in the model with random slopes and intercepts[9]. The dependent variable was the time to press the decision button (log_DT). We set 'Continue' as the reference condition for decision type as participants were informed in the task instructions that this decision was the preferred option in all videos. The reference category for condition was the SB condition.

We found a significant main effect of Condition (log_DT β = -0.17739, SE = 0.04960, t = 3.576, p = 0.001) together with a significant main effect of ‘Reverse’ decision type (log_DT β = 0.40, SE = 0.07, t = 5.41, p 0.001), a significant main effect of ‘Right’ decision type (log_DT β = -0.26, SE = 0.07, t = 3.89, p= 0.001) and an interaction effect between ‘Right’ decision type and HMD-360 condition (log_DT β = --0.13, SE = 0.06, t= -2.13, p = 0.03). There was no significant effect of ‘Left’ decision type (t = 1.12, p = 0.28) and no interaction between ‘Left’ x HMD-360 condition (t <1) or ‘Reverse’ x HMD-360 condition (t = -1.95, p = 0.06). Exponentiated Log DTs showed that decisions to Continue were made in the HMD-360 condition in 84% of the time

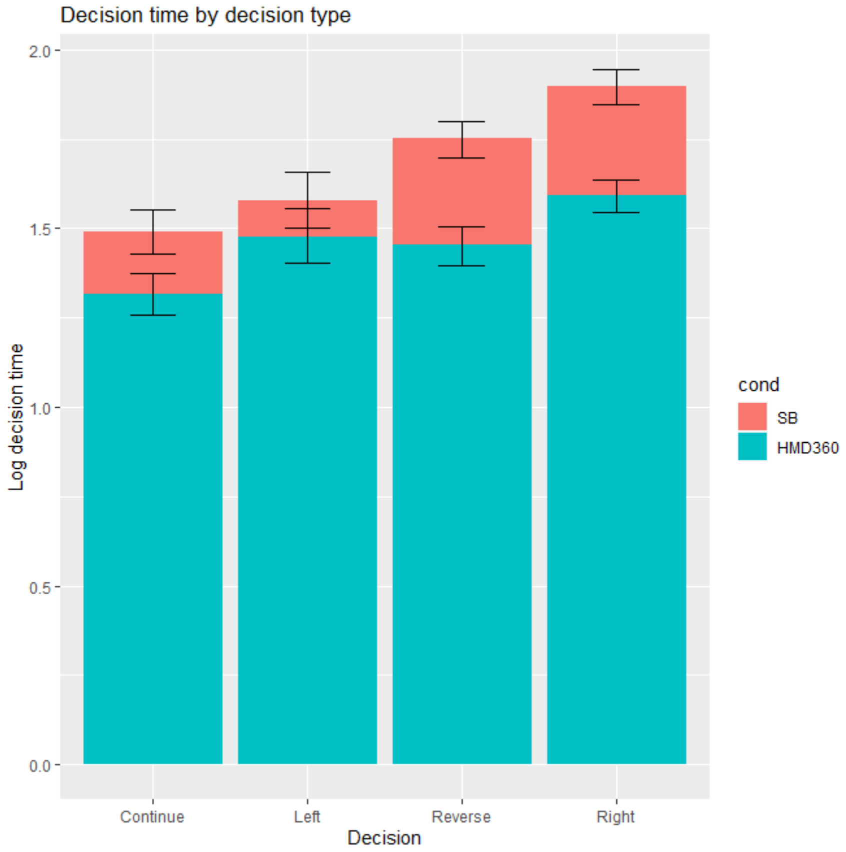


Figure 7 Decision time by answer type

---

[9] lmer(dt_log ~ cond* decision_type + (1+cond* decision_type |subj) + (1+cond* decision_type |video), data = combined_result)

required in the SB condition. In the SB condition, there was about a 49% increase in log time to make Reverse decisions and an increase of around 30% to make the decision to go Right compared to Continue decisions. The interaction effect between Right decisions and the HMD-360 condition showed that participants made decisions to go Right in 88% of the time taken in the SB condition when in HMD-360.

3.5.4 Exploratory analysis #4 – Effect of decision type on accuracy of decision

We also investigated whether some types of decisions in this study were more likely to result in accurate decisions than others, for example ‘Continue’ may have represented an easier judgement to make than ‘Reverse’ meaning a participant may be more likely to be correct for this decision. We specified a generalised linear mixed effects regression model[10] with instruction to treat the data as binomial, with Condition and Decision type as fixed effects and we added Subject and Video as random effects with by-participant random intercepts and slopes. The dependent variable was the accuracy of decision for each video. We kept Continue as the reference category for Decision type and SB as the reference category for Condition.

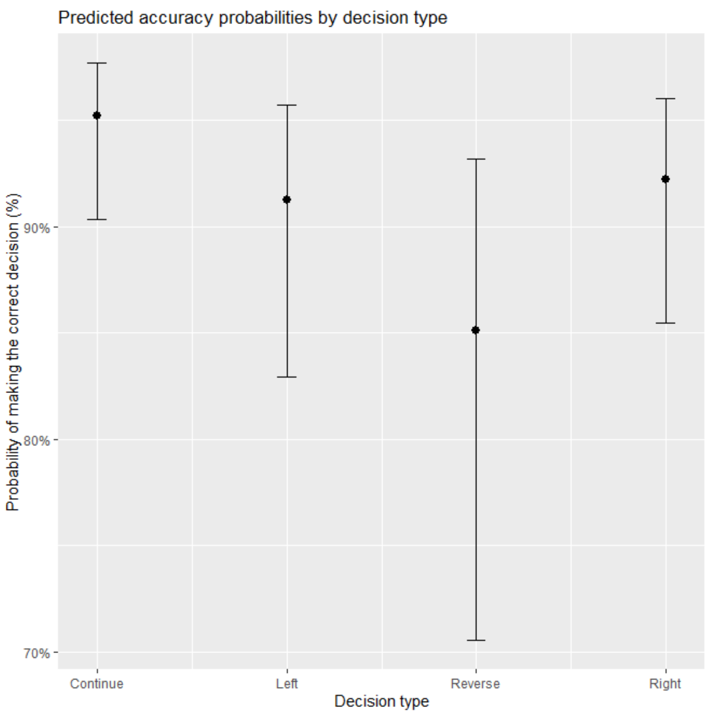


Figure 8 Probability of accurate decision by decision type

We found a significant main effect of Condition (β =0.73, SE = 0.269, t = 2.73, p = 0.01) and a significant main effect of ‘Right’ decisions (β =-1.05, SE = 0.44, t= -2.38, p = 0.02). We found no significant differences for ‘Left’ decisions (t = -1.02, p = 0.31) or ‘Reverse’ decisions (t = -1.34, p = 0.18).

Exponentiated model coefficients showed that the odds ratio of making a correct decision were twice as likely in the HMD-360 condition (2.08). Participants were over a third less likely (0.35) to make correct ‘Right’ decisions in the SB condition.

---

[10] *glmer(correct ~ cond + decision_type + (1 |subj) + (1|video), data = combined_results)*

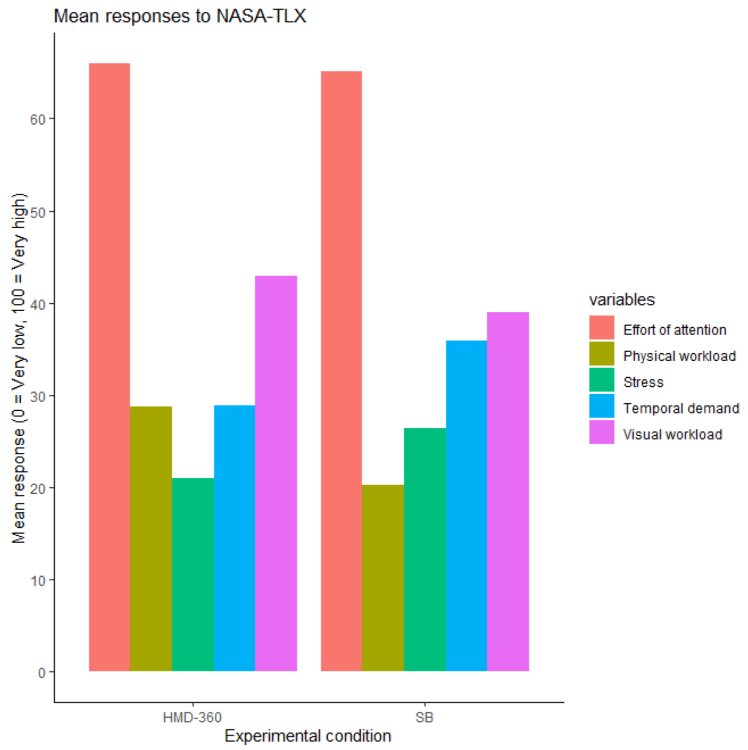


Figure 9 Mean response of Workload measures from NASA-TLX

3.6 Analysis of Workload measures

We asked participants to self-report visual, physical, and temporal workload pressures, as well as feelings of stress and attentional demands via the NASA TLX-R Workload measures (Hart, 2006). Responses were quantified on a sliding scale where 'very low' was 0 and 'very high' was 100.

Participants made verbal comments throughout the study that they found turning around in the HMD-360 condition to look at parts of the image behind them uncomfortable. Wilcoxon pairwise comparisons were carried out between NASA-TLX variables of effort of attention, physical workload, stress, temporal demand and visual workload with Bonferroni-adjusted alpha levels of 0.01 (0.05/5). Despite descriptive statistics suggesting that participants found the SB condition more stressful, but with lower physical workload, than the HMD-360 condition, there were no significant differences for any of the variables between conditions.

3.7 Analysis of Cyber sickness measures

Symptoms of Cyber sickness for discomfort, eye strain, headache, fatigue or dizziness were self-reported after each condition and answered on a separate Likert scale for each symptom. Data violated assumptions of a Chi-squared test due to the low numbers of observations for each participant (one per condition), a Fisher’s Exact test was used to determine if there was a significant association between SB and HMD-360 conditions for each cyber symptom.

There was no significant difference of self-reported symptoms of cyber-sickness between the two conditions for eye strain (p = 0.37), headache (p = 0.12), fatigue (p = 0.43) or dizziness (p = 0.12). There was a significant difference (p= 0.001) for discomfort with many participants reporting in the debrief comments that they found the headset in the HMD-360 condition extremely uncomfortable.

3.8 Analysis of Presence measures

Participants were asked 4 questions adapted from Schubert, Friedmann, & Regenbrecht's (2001) 13 item presence scale after each condition to measure presence. The question, *'How real did the virtual world seem to you '*, was scored between 1-100, with a lower score indicating 'completely real' and a higher score 'not real at all'. We compared participants' scores for SB and HMD-360 conditions using a Wilcoxon signed ranks test and found a significant difference (W(52) = 235, p = 0.0001) between perception of the virtual world in HMD-360 and SB conditions.

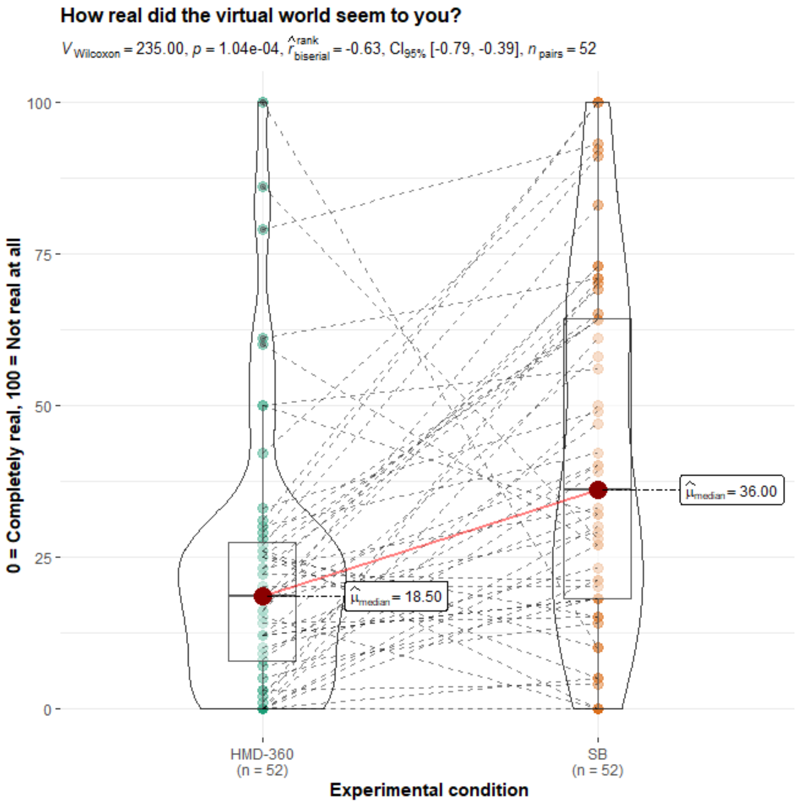


Figure 10 Plot of response to the question 'How real did the virtual world seem to you?'

However, Figure 10 shows that there were outliers in both conditions, which qualitative comments during testing that participant made aloud suggested that some participants were having difficulty interpreting the scale of the question.

The remaining three questions were scaled on a 5-point Likert response from "Strongly agree" to "Strongly disagree". A Fisher's Exact test was carried out on each question because of the low number of observations for each participant (1) for each condition. We found significant differences between conditions in the extent to which participants reported 'being there' (p = .003), 'feeling present' (p = 0.003) and 'being absorbed by the scene' (p = 0.001), with data skewed towards 'Agree' responses for the HMD-360 condition in all cases.

4.0 Eye tracking results

4.1 Eye tracking data collection

Eye movement data was collected in both conditions to make comparisons in distribution patterns between HMD-360 and SB. The measures between SB and HMD-360 presentation were as comparable as was feasible; moving the mouse to the edge of the screen in the monitor was the same position as the participant moving their eyes to the edge of the VR image without moving their head. The next section outlines the approach taken to transform the raw data in each condition into equivalent data for analysis.

4.2 Transformation of screen coordinates in SB condition

Throughout the SB condition, the Tobii Pro Nano screen-based eye-tracker collected 1) individual fixations on the screen/stimulus described by a single set of X, Y spatial coordinates and 2) participant actions defined in this study as the keypress and mouse clicks. We used fixations and mouse clicks to remap coordinates from screen to correspond with the 360° video image.

Tobii Pro Lab eye tracking software tracked participants' eye movements while they watched the videos on the monitor by recording the screen-based fixations coordinates. Participants can only see a 'window' of the original video image at any one time, although the image they are seeing completely fills their monitor screen (see Figure 11). The monitor screen pixel limits (1920 x 1080) are shown in Figure 11 superimposed onto the full video width limits (2048 x 1024) when they are positioned centrally to each other. If the participant was looking at the dead centre of the monitor (960, 540) this would correspond with the centre of the video (1024, 512).

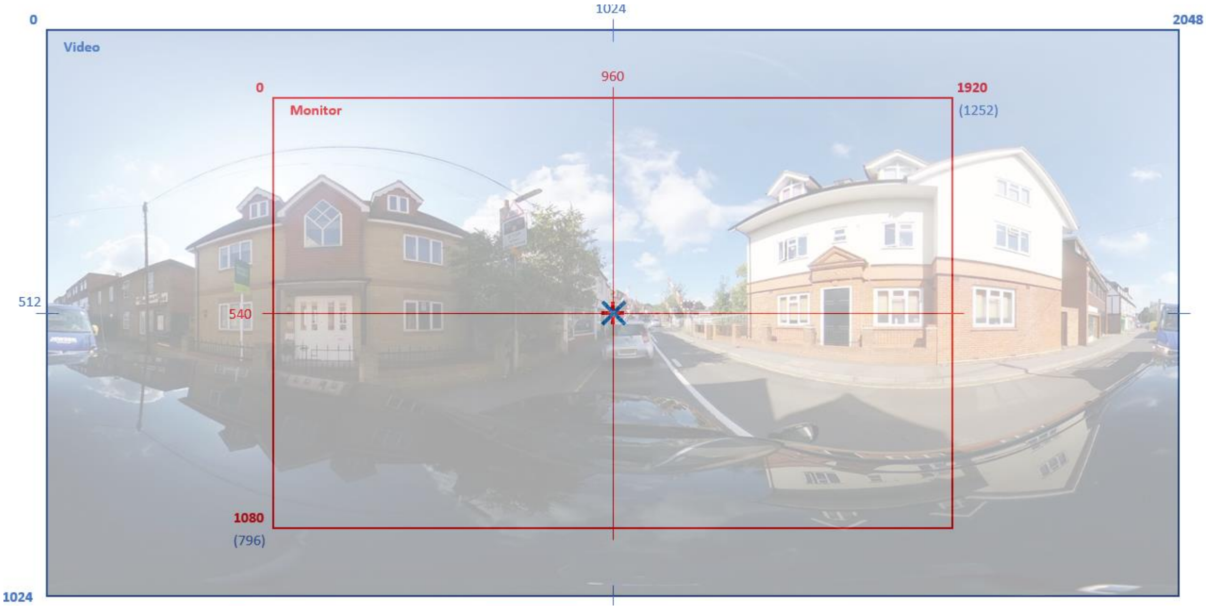


Figure 11 Monitor superimposed centrally on the full video

During the task, the participant uses the mouse by holding down the left-hand mouse button and dragging the image in the desired direction to view the portions of the image which are not in the viewing ‘window’. Tobii Pro Lab software records the mouse clicks (up, down) and X, Y positions of the mouse coordinates when the button is released, therefore eye movements and mouse clicks can be used to determine what portion of the image is presented in the monitor after the mouse drag. To determine the final fixation coordinates for the SB condition, we transformed the screen-based fixations into video image-based position. The next section outlines the calculations which were performed to reach the new coordinates for anlaysis.

4.2.1 Method of transformation of fixation coordinates

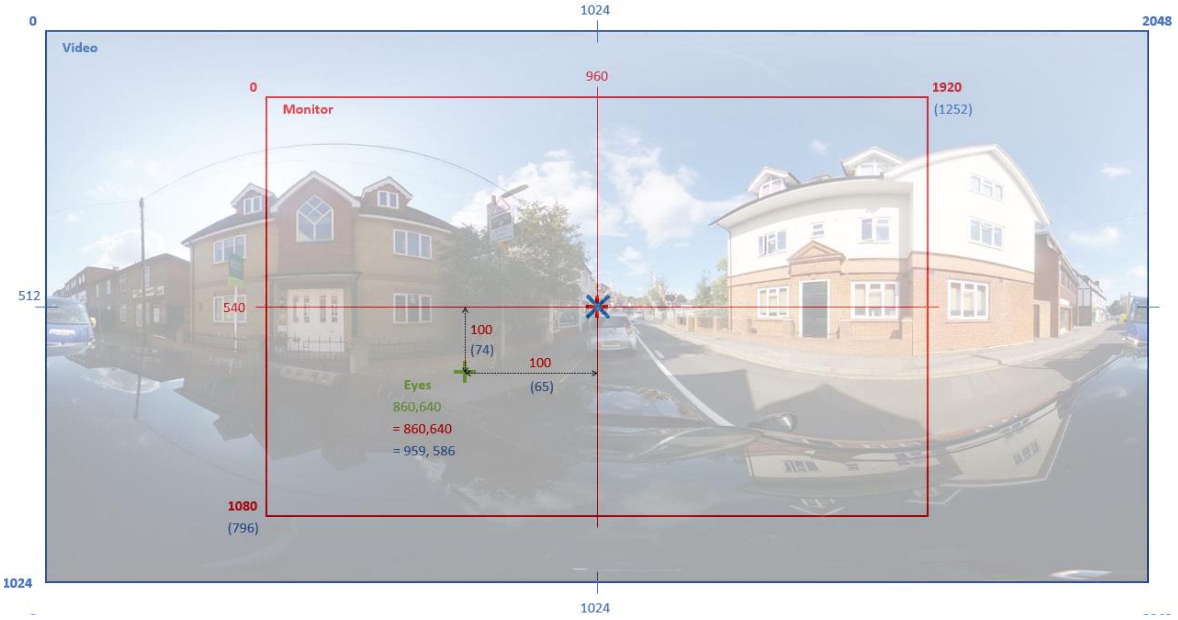


Figure 12 Monitor superimposed centrally on the full video with Eye positions

In Figure 12 the green cross indicates an eye position on the monitor. In this example, the eye fixation point is located on coordinates 860, 640 on the monitor. The position of the eyes in the video coordinates can then be calculated using the relative pixel difference between the video and monitor. In the X-axis, each 10 monitor pixels equals 6.52 video pixels (10*1252/1920) and in the Y-axis each 10 monitor pixels equals 7.37 video pixels (10*796/1024). Using these relative formulas, the eye position (shown in Figure 12) on the video can be calculated as 959, 586.

In Figure 13, the monitor screen has moved relative to the centre of the video as a result of the participant moving the mouse. In this example, the mouse drag was 80 monitor pixels right and 60 monitor pixels down, calculated as 52 video pixels left and 44 video pixels down, which positions the monitor centre on the video image at coordinates 1076, 596.

The eye fixation is located at 400,600 on the monitor. In the X-axis, this equates to 560 monitor pixels to the left of the monitor centre or 365 video pixels to the left of the coordinates calculated above of (1076, 596). In the Y-axis, this compares to 100 monitor pixels below the monitor centre or

74 video pixels to the below of the coordinates calculated above of (1076, 596). The eye fixation point on the video image in Figure 13 can therefore be calculated as coordinates 711, 630 (1076-365=711,556+74=630).

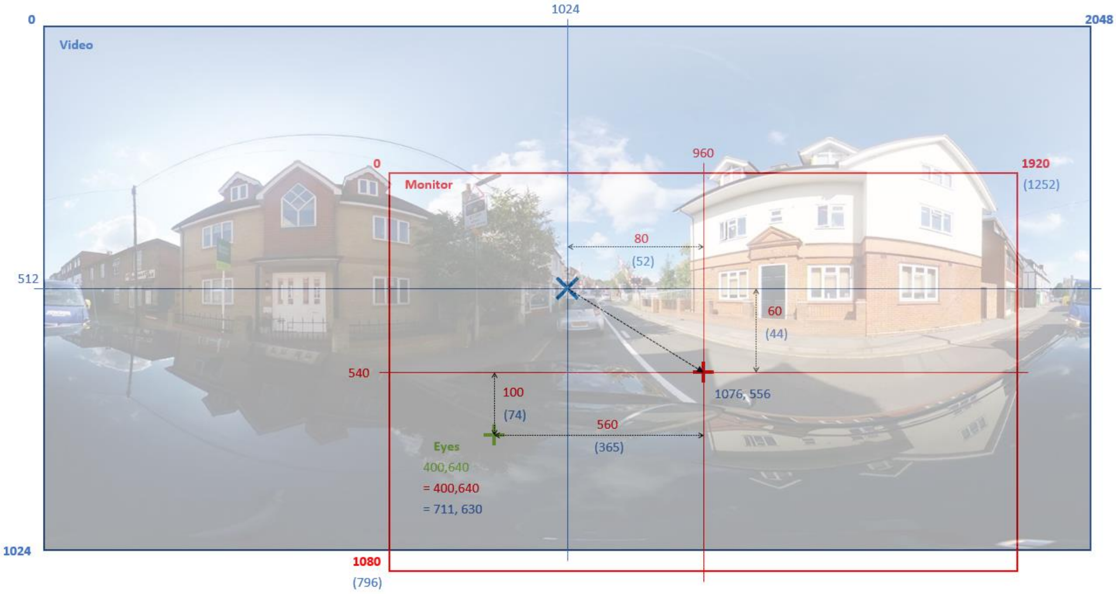


Figure 13 Monitor superimposed on the full video with initial mouse movement and Eye positions

As the participant continues to look around the video image to collect SA information before making their decision, Figure 14 shows another mouse drag relative to the previous [changed] monitor screen position in Figure 13. In this new example, the mouse change movement was 50 monitor pixels right and 42 monitor pixels up which is calculated as 32 video pixels right and 31 video pixels up, meaning the monitor centre is now located on the video image at 1108, 525.

Now the eye fixation is located at 760,450 on the monitor. In the X-axis, this compares to 200 monitor pixels to the left of the monitor centre or 130 video pixels to the left of the coordinates calculated above of (1108, 525). In the Y-axis this equates to 84 monitor pixels above the monitor centre or 74 video pixels above the coordinates calculated above of (1108, 525). The new eye fixation point on the video can therefore be calculated as coordinates 978, 463 (1108-130=978,525-62=463).

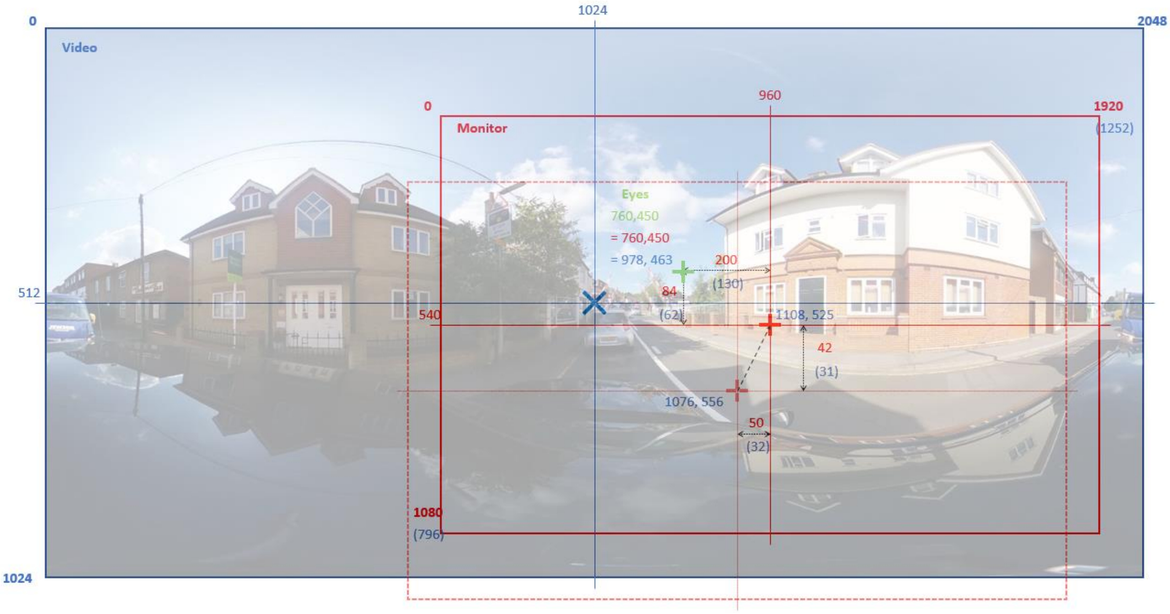


Figure 14 Monitor superimposed on the full video with subsequent mouse movement and Eye positions

Finally, due to the 360° nature of the videos, if a participant moves keeps moving the mouse, they can end up back at the same place on the video, as the image is cylindrical (e.g., X= 370° becomes X=10°). In this event, we corrected the transformation by increasing or decreasing the new coordinate value by 2048 (the full width of the video image).

The same process of transformation described above was used throughout each trial across all participants in the dataset for the SB condition. The transformed X,Y coordinates were then used for the final analysis described in section 4.7.

4.3 Extracting gaze coordinates in HMD-360 condition

The input in both conditions is a 2D representation of a 360° world as a spherical equirectangular-panoramic image. In the HMD-360 condition, it is wrapped around the participant using the headset display whereas in the SB condition it is presented on screen. In the HMD-360 condition, Tobii Pro Lab presents the video image stretched in the headset with pixel limits of X (0, 4096) and Y (0, 2048). These are exactly two-fold of the original video resolution of 2048 x 1024. In the HMD-360 condition, the Tobii Pro Lab software recorded fixations and mapped them from the display to coordinates in the 2D equirectangular- panoramic image. To map the fixation coordinates from the HMD-360 headset data onto the video image X,Y coordinates, we divided them by two and used these new gaze coordinates for the analysis.

4.4 Creation of areas of interest in the driving videos

In the present study, we defined four areas of interest (AOIs) across all the videos on which to analyse fixations. To correspond with the four decision options on the keypad for the experimental task (continue forward [8], go left [4], go right [6] and reverse [2]), we defined AOIs as a tetradic division of the video image (Centre/Left/Right/Back). We split the full video image into 4 equal 'slices' from left to right at 512 pixels per slice to represent the 360° spherical video image wrapping around the participant in the headset (see Figure 15).

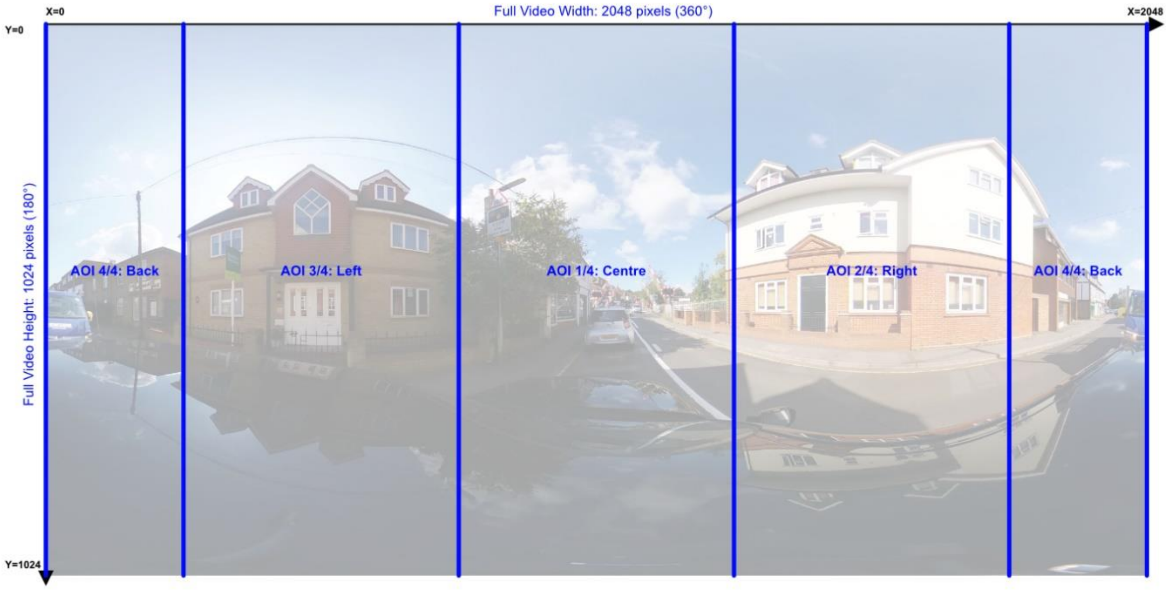


Figure 15 Defined areas of interest (AOIs) in the video images

One half of the 'Back' AOI is 1/8 image on the far left of the video image, and the other half of the 'Back' AOI is 1/8 of the image on the far-right hand-side of the video image of the 2D equirectangular projection used as input. The 'Left', 'Centre' and 'Right' AOIs were each allocated a continuous quarter of the equirectangular input video. We used the full range of Y coordinate from 0- 1024 for each AOI.

We defined each AOI label as a category of interest in R Studio and recorded every calculated fixation point X coordinate which fell within that AOI. The Back1 and Back 2 AOIs were combined after AOI analysis into a single AOI called 'Back'. We removed all fixation data for the first 5 seconds of each trial for all participants for both conditions to account for the fixation cross at the start of each video, so no AOI fixations were collected during that time. See Table 6 for the X coordinates for each AOI.

Table 6 Category labels and x, y coordinates for AOIs
Nb Back1 and Back 2 are combined after AOI analysis into a single AOI called "Back"

Area of Interest	Fixation X coordinates	Fixation Y coordinates
Back 1	0-256	0, 1024
Left	256-768	0, 1024
Centre	768-1280	0, 1024
Right	1280-1792	0, 1024
Back 2	1792-2048	0, 1024

4.5 Eye tracking dependent measures

We predicted that there would be differences in gaze distributions between the two conditions shown by patterns of fixations, for example a participant might have been looking central AOI more frequently in HMD-360 condition compared to SB condition. To allow for differences between participants in the length of trials (as some participants may look at the video for more or less time before making their decision), we measured fixation duration as the average percentage of time spent looking in each of the four AOIs ('Left', 'Centre', 'Right' and 'Back') for both conditions (HMD-360 and SB) for each participant.

4.6 Data cleaning decisions

We kept the same criterion for a trial as the behavioural analyses (see section 3.1). Participants had a total of 30 trials in each condition and one keypress (2,4,6 or 8) was recorded as a response to each trial. The time between the start of the video and the keypress decision was recorded as decision time (DT) with 5 seconds subtracted for the duration of the fixation cross at the start of each video. We used the X fixations which had been transformed by the processes described in sections 3.1.4 and 3.1.5.

We removed trials which had eye tracking data loss of >20% and four participants were removed from the dataset due to data loss in more than 50% of trials. Previously, five participants had been removed from the dataset in the behavioural analyses outlined in section 3.0, meaning the eye tracking analysis was conducted on a total of 43 participants.

4.7 Analysis of eye tracking results

Descriptive statistics (see Table 7) showed that there were differences in the duration of fixations across AOIs ('Left', 'Centre', 'Right' and 'Back') between the two conditions (HMD-360/SB).

Table 7 Descriptive statistics to show differences in the duration of fixations across AOIs in each condition

	Condition: SB			Condition: HMD-360		
AOI	*M*	*M* 95% CI [LL, UL]	*SD*	*M*	*M* 95% CI [LL, UL]	*SD*
Left	17.38	[15.44, 19.32]	6.30	5.52	[4.49, 6.54]	3.32
Centre	66.99	[64.28, 69.69]	8.79	84.45	[82.26, 86.64]	7.12
Right	10.73	[9.86, 11.61]	2.84	7.56	[6.47, 8.65]	3.53
Back	4.90	[4.18, 5.62]	2.34	2.47	[1.55, 3.39]	2.99

A 2x4 repeated measures ANOVA was carried out with factors of Condition (HMD-360 and SB) and AOI (‘Left’, ‘Centre’, ‘Right’ and ‘Back’). There was no possible main effect of Condition as the fixation percentages sum to 100% in both conditions, so they could never differ. There was a significant main effect of AOI (F(1, 42) = 1952.456 , p < 0.0001) and a significant interaction between Condition and AOI (F(1, 42) = 160.956, p < 0.0001), such that participants looked in different AOIs for longer (or shorter) durations depending on which condition they were in (see Figure 16).

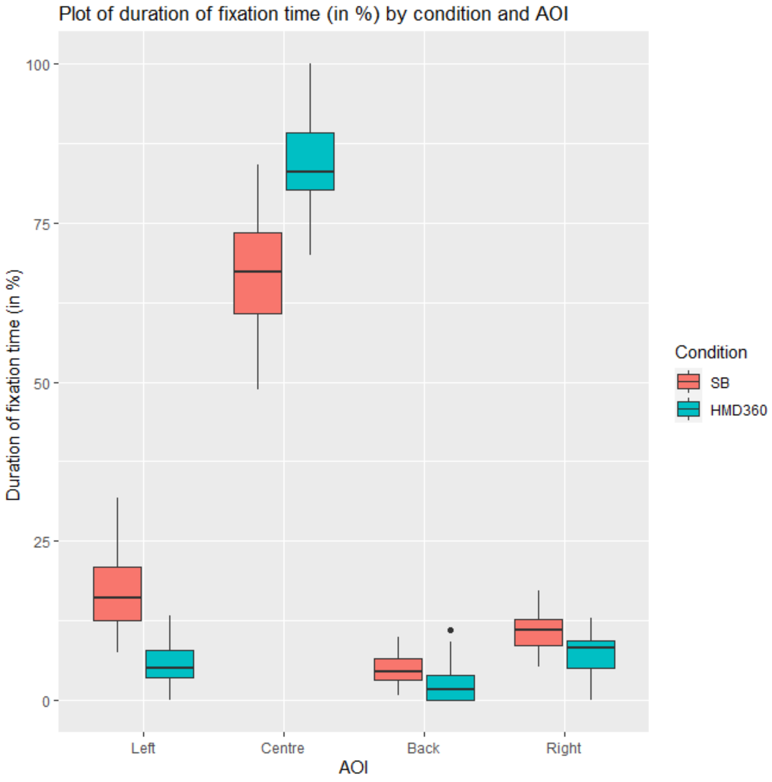


Figure 16 Proportion of time spent in each AOI by condition

Post-hoc comparisons were performed on the effect of AOI on fixation duration (%) for each condition with Bonferroni corrections and showed that all comparisons were significantly different at p = 0.001 (except for HMD-360: Left/Right which was significant at p = 0.05 and HMD-360: Left/Back was significant at p = 0.01) (see Table 8 for test statistics and significance levels for all interaction effects.).

There were more ‘Centre’ AOI fixations overall than other AOIs and a significantly higher percentage duration of ‘Centre’ AOI fixations in the HMD-360 condition. A significantly higher percentage of time was spent looking in the ‘Left’, ‘Right’ and ‘Back’ AOIs in the SB condition than in the HMD-360 condition, with ‘Left’ AOIs being significantly higher than ‘Right’ and ‘Back’ AOIs in the SB condition.

Participants spent the least proportion of time looking in ‘Back’ AOIs in both conditions, but a significantly lower proportion of time looking in ‘Back’ AOIs in the HMD-360 condition compared to the SB condition. Participants also spent a significantly higher proportion of time looking in ‘Centre’ AOI’s compared to ‘Back’ AOIs in the HMD-360 condition.

Table 8 Post-hoc Bonferroni corrected comparisons for the effect of Condition on fixation duration % at every AOI.

AOI	Condition	Condition	statistic	df	p	Bonferroni .adj	p.adj.signif
Left	SB	HMD360	13.35297	42	1.04E-16	1.04E-16	****
Centre	SB	HMD360	-14.1672	42	1.34E-17	1.34E-17	****
Back	SB	HMD360	5.771194	42	8.47E-07	8.47E-07	****
Right	SB	HMD360	5.504884	42	2.04E-06	2.04E-06	****

5.0 Discussion

5.1 Comparisons with other studies

In this paper, we investigated remote supervision using a choice decision task to assess the effect of two formats of presentation: HMD-360 and SB. Our results showed that on average, decisions were made in the HMD-360 condition in 80% of the time required in the SB condition with decisions getting slightly quicker for each trial over 30 trials and participants more likely to give correct answers when videos were presented in the HMD-360 condition. These findings contradict previous research (Georg et al., 2020) which reported no significant findings between different display types on SA decision making. We also found evidence from eye tracking data that there were significant differences in the duration of fixations in different parts of the video image depending on whether it was presented in SB or HMD-360 format.

Unlike some studies discussed previously, which compared HMD with a monitor-only 360° setup for controlling a teleoperated vehicle and found no significant differences in reported sense of presence between the two setups, in our study, we found a significant effect of the HMD-360 condition in the extent to which participants reported 'being there', 'feeling present' and 'being absorbed by the scene'. The higher ratings of the sense of presence in the HMD-360 (vs. the SB 360) condition in our study corresponded with higher performance times and accuracy of decisions in that condition. This supports earlier studies that have reported positive relationships between a higher sense of presence and task performance (Georg et al., 2020; Witmer & Singer, 1998). In keeping with findings by Georg et al. (2020) who observed no significant difference between monitor display/HMD-360 and workload, we also found no difference of perceived workload between the two conditions.

Other research (Weidner et al., 2017) into the use of HMD-360 display modes on driving performance found that HMD-360 significantly increased simulator sickness whereas participants in our study reported no cyber sickness symptoms in either condition. The estimated eye tracking delay in our study was about 50 ms, and end-to-end latency was about 79 ms but participants viewed the remote scenes from a similar position as an RO so, although there were dynamic movements of other cars and pedestrians in the scenario to assess, the participant themselves were not in motion and thus did not have conflicting vestibular information which is probably why there were fewer experiences of cyber sickness.

Although research has investigated whether HMD-360 improves operator performance (see for example, Grabowski, Jankowski, & Wodzyński, 2021; Monteiro et al., 2020; Zhang et al., 2019) or SA of hazards (Crundall et al., 2021), no studies, to our knowledge, have investigated the effect of the particular type of edge case that is encountered on the time it takes to make SA decisions. We found evidence from exploratory analyses that the type of edge case shown in the video influenced the time taken to make decisions. Participants, on average, were slower at making decisions for videos depicting dead ends or closed gates. These findings suggest that participants found the types of scenarios presented by these videos more challenging. This could be, at least in part, due to the uncertainty that these scenarios present. For example, in video 174, there was a security guard

Figure 17 Video 174 nb. the security guard is to the right of the vehicle

visible in the scene which may have led some participants to decide to wait to see if they would open the gate, which would then have enabled a 'continue' rather than 'reverse' decision (see Figure 17).

Classifications involving dead ends also revealed a decision time delay. For example, Video 162 had a footpath to the left of the vehicle which would not be big enough to drive a vehicle through but still needed to be evaluated, also resulting in a time consequence. These empirical findings reveal, predictably, that edge cases which are ambiguous in nature will require more time for ROs to develop SA before they can come to a decision.

Decisions in response to videos which required highway code comprehension e.g., passing parked cars on the side of the road by moving into the opposite lane, or responding to objects in the road e.g. a scooter, or following road way signs and signals such as right turn only, took less time than was taken in other videos. These findings indicate that participants may have found these types of edge cases less demanding to develop the necessary SA to come to a decision as they were more straight forward. We also investigated whether the type of edge case scenarios would impact on participants' ability to make correct decisions in response to the videos. Although participants took longer to analyse videos depicting dead ends, exploratory analyses showed that they were over twice as likely to make the correct decision, showing that the extra time taken to develop adequate SA was wise in these instances.

In Paper 3, we conducted an exploratory factor analysis on 48 items designed to measure SA of driving videos and found four underlying constructs which we termed ‘spatial and environmental awareness’ (Factor 1), ‘anticipatory hazards’ (Factor 2), ‘dynamic driving actions’ (Factor 3), and ‘other road users’ (Factor 4). In this study, we didn’t directly ask participants SA questions, nonetheless it is interesting to revisit the effect of video type considering these factor labels. Video Type 5 [Highway code] and Type 7 [Signage] corroborate the elements identified in Factor 1 relating to perception and comprehension of signs in the remote scene, demonstrating that these are highly relevant to the construction of remote SA. Type 2 [Dead ends], Type 4 [Gates] and Type 6 [Object in the road] could also fall under the construct of ‘anticipatory hazards’ in Factor 2. Trials involving Type 2 [Dead ends] and Type 4 [Gates] videos took participants significantly longer to make their decision which provides further support that evaluation of this underlying construct is important when building SA. Although the experimental task in this study did not require any direct driving control, participants had to project what the future on-going direction of travel would be which implicitly supports the dynamic driving actions outlined in Factor 3. Finally, the videos in this study were designed to limit interaction with other road users, for example pedestrians, so Factor 4 is not directly applicable to this task, yet the influence on decision times of the presence of the security guard in one video, discussed above, demonstrates that 'other road users' are still likely to be factored into supervisory decision making confirming the importance of this dimension in the construction of SA from remote video feeds. Although still in the initial stages of development with much work to be done in this area of research, there are indications from our findings in this study that, in the future, for any instrument design of a new measure of remote SA, consideration of these types of edge cases would be an important inclusion.

Eye tracking data analyses demonstrated differences in the duration of fixations in distinct sections of the video image depending on whether it was presented in SB or HMD-360 format. All the videos started with a fixation cross, directing attention towards the forward-facing view so, unsurprisingly, eye movement data showed more 'Centre' AOI fixations overall than any other AOIs in both conditions. Even in the light of studies of visual perception that have found a systematic bias to look at this location first (Bindemann, 2010), in our study there was a disproportionate amount of 'Centre' AOI fixations in comparison to other areas of the video image since, in the HMD-360 condition, if a participant looked left and found no relevant information in that AOI, they were required to move back through the centre AOI to look right. Even in the SB condition, where participants could continue scrolling left to get back the centre view (due the spherical nature of the 360° image), most of the time participants did not do this and scrolled to the right, through the centre.

However, the significantly higher percentage duration of 'Centre' AOI fixations in the HMD-360 condition could represent a more parsimonious search strategy than in the SB condition, as participants only had to move their head slightly to the left or right to access SA information not directly in front of them and may have been able to access peripheral information in the scene more quickly. To look around the scene in the SB condition, participants had to drag the image with the mouse, producing less subtle movements and generating wider sampling of different AOIs in the image. Research has found that mouse movements commonly start with fast but unfocused movement, then slow down and gain accuracy: but when combined with high cognitive load conditions result in longer task duration and more direction changes (Grimes & Valacich, 2015). Increases in DT in the SB condition to make Reverse and Right decisions may be a consequence of having to spend additional time scrolling around the scene to gather SA information to come to that decision choice, compared to glancing right or quickly twisting around in the chair to look behind them in the HMD-360 condition.

The lowest proportion of time spent in fixating in ‘Back’ AOIs in both conditions was surprising, particularly as 30% of correct decisions for the 60 videos were ‘Reverse’. This could be as once participants had eliminated ‘Continue’, ‘Left’ and ‘Right’ decisions in response to the task it was unnecessary to continue to search in the video for the answer as only ‘Reverse’ remained as a choice. However, ‘Reverse’ decisions were significantly slower in the HMD-360 condition than in the SB condition, which could reflect the unwillingness of participants to twist around in the static chair to fully access the parts of the 360° image behind them because it was uncomfortable. This highlights the possible advantage of screen-based presentation to assist the development of SA in remote feeds as the back parts of the image were more easily accessible via mouse drag.

Previous research (Buscher et al., 2009) indicated that goal-directed information searches on computer screens would be focused on the top left of the screen. We found a greater proportion of ‘Left’ AOI distributions in the SB condition than HMD-360 although we did not measure whether they were specifically looking in the top section of the AOI as we were only interested in horizontal X-axis coordinates when defining AOIs and did not analyse vertical coordinates. Interestingly, in the SB condition, the outcome of a mouse movement to the right shifted the viewing window on-screen to the left, although the participant’s intention may have been to look right of the image. However, the significantly higher proportion of ‘Left’ AOI fixations compared to ‘Right’ AOI and ‘Back’ AOIs across many trials in the SB condition point toward a deliberate strategy rather an unintended error. In comparison, the proportion of ‘Right’ AOI fixations were the second highest of AOI fixations in the HMD-360 condition (after ‘Centre’ AOI fixations).

Participants looked left more frequently in the SB condition and looked right more frequently in the HMD-360 condition, which is an important behavioural discrepancy between the two conditions in the context of the RO supervisory task they were performing. It is possible that participant’s higher self-reported sense of presence in the HMD-360 condition meant they felt more embodied in the vehicle, resulting in more naturalistic viewing behaviour which mirrored behaviour in situ. As our sample was a UK left hand traffic population, they may have favoured ‘Right’ AOIs in the HMD-360 condition as per usual driving protocols of checking right, left, right before carrying out any driving manoeuvres. Additionally, this may explain the interaction effect between Right decisions: HMD-360 as they had an arguably better chance of correctly interpretating SA information if they looked in that direction, unlike the SB condition who were predominantly looking in the ‘Left’ AOI. This means that visual search strategies may be different when remote video feeds are presented via monitor display compared to HMD which, in turn may impact on the time it takes to come to a decision in a supervisory task and the accuracy of that decision.

5.2 Limitations and modifications

One limitation in this study was that some participants lacked proficiency in using the mouse in the SB condition to click through the 360 videos on the monitor. This could have lengthened DTs in the SB condition for them compared to the HMD-360 condition which did not require the use of a mouse. We allowed participants as much practice as they required to build necessary skills to perform the task competently in both conditions, but it may have been advisable to pre-screen participants for how much time they habitually spent using a computer and mouse set-up before starting the experiment. Future studies that require screen-based interaction should only include subjects who are experienced in mouse control as a sample restriction.

Very few participants in this study had any experience in HMD-360: only 3% of participants in this study owned an HMD-360 headset with most (56%) having never used a headset before, and the others only one or two times (41%), so this would probably have enhanced the learning effects we saw on a trial-by-trial basis in the HMD-360 condition by comparison with experienced HMD-360 users. Again, using pre-screening protocols to match participants by experience would be sensible in future studies.

Perhaps due to their inexperience, many participants complained that they found the headset in the HMD-360 condition extremely uncomfortable: one participant commented in the debrief,

> *"Wearing the HMD-360 headset put pressure on the nose and made you not feel like looking behind you as the weight of the headset made it not an easy task."* Participant 16.

Several other participants echoed this sentiment, commenting that turning to look behind them would have been "*easier in a swivel chair*" (Participant 38) rather than the fixed chair we used in the lab set-up. This option was rejected during piloting as it increased dizziness. Weidner et al. (2017) placed participants in a mock up car to increase their sense of presence, so we could have used this tactic to provide a more realistic feeling of sitting in a car where you would access information behind you by glancing over your shoulder rather than spinning in a chair. However, in future it is more likely that access to visual information from behind the AV will be via interface control (although we note that this may in turn reduce the sense of ‘presence') and more recent HMD eye trackers are significantly lighter than the HTC Vive model we used. Careful testing to judge the relative benefits of one approach over the other will be necessary to identify the optimal solution.

There may have been difficulty viewing some of the information in the videos in both conditions. To avoid software crashes in the HMD-360 condition, the resolution was down sampled for videos in

Figure 18 Video 232 showing sub-optimal lighting conditions behind the vehicle nb. the view behind the vehicle is represented to the right and left of the 360 image.

both conditions to 2048x1024. This could have affected decision times by reducing the ability to inspect road signs and other objects in detail. However, previous studies (Georg et al., 2020). which manipulated the quality of presentation in monitor and HMD-360 based presentation did not find that low quality settings resulted in fewer road users or traffic signs being detected compared to higher settings In this study, all videos were viewed equally either on the monitor or in the HMD-360 headset over the course of the study and participants made correct SA judgements over 90% of the time in both conditions, which suggests that the lowered resolution did not impede the ability to make decisions in our study, although we recognise that any decisions to reduce resolution further may negatively affect SA performance.

Also, although consistent across conditions, the lighting conditions in some of the videos were sub-optimal, for example, in video 232, the road behind the vehicle is very dark due to overhanging trees in the background (shown in Figure 18), which could have added time to participants' decisions while they tried to discern if there was a possible route in that direction. However, ROs are also likely to experience variations in the quality of the video feed due to lighting conditions at the AVs location, thus increasing the experimental realism of our videos in representing the range of scenarios that an RO may encounter.

The differences we found between decision times for the SB condition and the distribution of attention in AOIs compared to the HMD-360 condition may have been due to the method used to control the video feed. Although monitor presentation is typical in most industry-based remote operations, adjusting the visual input using a mouse, as participants did in our study, is less regular. Commonly a joystick, game controller or steering controls are used to control the AV and visual feeds from the remote scene are altered using allocated buttons on the control device to shift between cameras fixed to the AV, rather than moving around a 360° image. However, novel interface prototypes for remote operation have proposed stitching together images from the cameras placed around the AV as a 360° image, so this research can help to inform designers of the complexity of accessing information in 360° in a naturalistic way.

5.3 Future research and recommendations

Although we are still likely to be many years away from sharing public highways with AVs, the AV industry continues to grow. The transport industry urgently needs to consider what mode of presentation should be utilised to deliver relevant information from the remote scene to ROs in order to enable them to build and maintain remote SA as effectively as possible, as well as investigating how much time it is likely to take to make decisions based on second-hand information from a remote scene. It is recommended that companies offering remote supervision of AVs explore the potential of HMD-360 presentation to enhance operators' remote SA. The results of this study further suggest that it would be most appropriate to limit HMD-360 to shorter periods of operation to reduce negative workload issues. This indicates that HMD-360 may be more suitable for operators carrying out a supervisory role in fleet operations, where the nature of the engagement is likely to be briefer, than for operators engaging in extended periods of teleoperation.

The task of gaining and maintaining SA for ROs depends on the different types of vehicles they may operate, for example cars, trucks or forklifts, the RO use case, either direct remote control or supervisory tasks and the segment of the transportation industry that they are working in. The findings of this study are relatable to all these domains, as we have demonstrated that there are critical behavioural and eye movement differences when using screen-based video feed presentation and other modes of presentation to provide ROs with visual feeds from the remote scene. There are also significant implications of our results for remotely operating on-road vehicles in different countries requiring switches between left- and right-hand driving for example driving in the UK and Europe. Wide adoption of remote operation in many industries has the potential to open up the labour market by investing in a global workforce who can remotely operate any vehicle in any location from anywhere in the world but careful empirical testing using eye tracking software to understand where ROs look in the remote scene to gather information and designing interfaces accordingly will be an important direction for future transportation research.

Little is known about the effectiveness of different modes of presentation of video feeds from remote scenes in terms of their ability to support a remote operator in gaining and maintaining sufficient SA to make accurate decisions. This paper contributes to the progression of automated transport solutions by exploring and empirically testing the effects of HMD-360 compared to standard screen-based presentation in an applied task that is directly applicable to the remote supervision of AVs. The results of this study show that humans also take time to interpret these unusual events, but participants consistently responded with a high degree of accuracy, showing that unlike AVs, we are capable of interpreting novel scenarios. This demonstrates that human-machine interaction is likely to remain a critical feature in the automated vehicles of the future (Law Commission of England and Wales, 2022).

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*(23), 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with some simple guidelines. *Journal of Cognition*, *2*(1), 1–38.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

Buscher, G., Cutrell, E., Morris, M. R., & Way, O. M. (2009). *What Do You See When You ' re Surfing ? Using Eye Tracking to Predict Salient Regions of Web Pages*. 21–30.

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human-Computer Interaction*, *36*(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

de Winter, J. C. F., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology and Work*, *21*(1), 99–111. https://doi.org/10.1007/s10111-018-0527-6

Georg, J., Feiler, J., Diermeyer, F., & Lienkamp, M. (2018). *Teleoperated Driving , a Key Technology for Automated Driving ? Comparison of Actual Test Drives with a Head Mounted Display and Conventional Monitors **. 3403–3408.

Georg, J. M., Putz, E., & Diermeyer, F. (2020). Longtime Effects of Videoquality, Videocanvases and Displays on Situation Awareness during Teleoperation of Automated Vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *2020-Octob*, 248–255. https://doi.org/10.1109/SMC42975.2020.9283364

Grabowski, A., Jankowski, J., & Wodzyński, M. (2021). Teleoperated mobile robot with two arms: the influence of a human-machine interface, VR training and operator age. *International Journal of Human Computer Studies*, *156*(December 2020). https://doi.org/10.1016/j.ijhcs.2021.102707

Grimes, G. M., & Valacich, J. S. (2015). Mind over mouse: The effect of cognitive load on mouse movement behavior. *2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015*, (October).

Hart, S. G. (2006). NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER. *Proceedings of The Human Factors and Ergonomics Society 50th Annual Meeting*, 904–908.

Kennedy, Lane, Berbaum, & L. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3(3), 203-220*.

Law Commission of England and Wales. (2022). *Automated Vehicles : joint report*.

Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, *23*(6), 736–757. https://doi.org/10.1080/13506285.2015.1079583

Monteiro, D., Liang, H. N., Wang, J., Chen, H., & Baghaei, N. (2020). An In-Depth Exploration of the Effect of 2D/3D Views and Controller Types on First Person Shooter Games in Virtual Reality. *Proceedings - 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2020*, (1), 713–724. https://doi.org/10.1109/ISMAR50242.2020.00102

Mutzenich, C., Durant, S., Helman, S., & Dalton, P. (2021). Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles. *Cognitive Research: Principles and Implications*, *6*(1). https://doi.org/10.1186/s41235-021-00271-8

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Shahar, A., Alberti, C. F., Clarke, D., & Crundall, D. (2010). Hazard perception as a function of target location and the field of view. *Accident Analysis and Prevention*, *42*(6), 1577–1584. https://doi.org/10.1016/j.aap.2010.03.016

Simpson, B. D., Bolia, R. S., & Draper, M. H. (2013). Spatial audio display concepts supporting situation awareness for operators of unmanned aerial vehicles. In *Human Performance, Situation Awareness, and Automation: Current Research and Trends HPSAA II, Volumes I and II* (p. 61).

van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment. *International Journal of Aviation Psychology*, *22*(1), 78–95. https://doi.org/10.1080/10508414.2012.635129

Weidner, F., Hoesch, A., Poeschl, S., & Broll, W. (2017). Comparing VR and non-VR driving simulations: An experimental user study. *Proceedings - IEEE Virtual Reality*, 281–282. https://doi.org/10.1109/VR.2017.7892286

Wilkie, R. M., Kountouriotis, G. K., Merat, N., & Wann, J. P. (2010). Using vision to control locomotion: Looking where you want to go. *Experimental Brain Research*, *204*(4), 539–547. https://doi.org/10.1007/s00221-010-2321-4

Zhang, G., Minakata, K., & Hansen, J. P. (2019). *Enabling Real-Time Measurement of Situation Awareness in Robot Teleoperation with a Head-mounted Display*. 50–52.

Appendix

Full list of video numbers, description of edge case the video depicts, playlist details and correct answer and keypress with scenario classification.

Video number	Playlist	Correct answer	Correct keypress	Description of the video edge case	Scenario classification
206	A	Continue	8	Construction works	Construction works
216	A	Continue	8	Construction works	Construction works
194	A	Continue	8	Railway	Highway code comprehension
203	A	Continue	8	Parked cars	Highway code comprehension
212	A	Continue	8	Parked cars	Highway code comprehension
231	A	Continue	8	Road turns	Highway code comprehension
236	A	Continue	8	Object in road	Object in road
237	A	Continue	8	Object in road	Object in road
238	A	Continue	8	Object in road	Object in road
172	A	Left	4	Cones blocking	Diverted route
164	A	Left	4	Signs/signals	Signs/signals
169	A	Left	4	Signs/signals	Signs/signals
183	A	Reverse	2	Dead end	Dead end
185	A	Reverse	2	Dead end	Dead end
210	A	Reverse	2	Dead end	Dead end
219	A	Reverse	2	Dead end	Dead end
222	A	Reverse	2	Dead end	Dead end
226	A	Reverse	2	Dead end	Dead end
186	A	Reverse	2	Cones blocking	Diverted route
174	A	Reverse	2	Gates	Closed gates
175	A	Reverse	2	Gates	Closed gates
200	A	Reverse	2	Gates	Closed gates
211	A	Reverse	2	Gates	Closed gates
240	A	Right	6	Construction works	Construction works
163	A	Right	6	Cones blocking	Diverted route
214	A	Right	6	Cones blocking	Diverted route
218	A	Right	6	Gates	Closed gates
209	A	Right	6	Road turns	Highway code comprehension
228	A	Right	6	Road turns	Highway code comprehension
213	A	Right	6	Signs/signals	Signs/signals
178	B	Continue	8	Construction works	Construction works
176	B	Continue	8	Parked cars	Highway code comprehension
177	B	Continue	8	Parked cars	Highway code comprehension
193	B	Continue	8	Railway	Highway code comprehension
195	B	Continue	8	Railway	Highway code comprehension
201	B	Continue	8	Parked cars	Highway code comprehension
205	B	Continue	8	Parked cars	Highway code comprehension
220	B	Continue	8	Object in road	Object in road
198	B	Continue	8	Signs/signals	Signs/signals
234	B	Left	4	Construction works	Construction works
232	B	Left	4	Cones blocking	Diverted route
196	B	Left	4	Signs/signals	Signs/signals
208	B	Left	4	Signs/signals	Signs/signals
170	B	Reverse	2	Construction works	Construction works
173	B	Reverse	2	Construction works	Construction works
159	B	Reverse	2	Dead end	Dead end
162	B	Reverse	2	Dead end	Dead end
221	B	Reverse	2	Dead end	Dead end
165	B	Reverse	2	Gates	Closed gates
171	B	Reverse	2	Signs/signals	Signs/signals
242	B	Reverse	2	Signs/signals	Signs/signals
157	B	Right	6	Construction works	Construction works
188	B	Right	6	Construction works	Construction works
207	B	Right	6	Construction works	Construction works
158	B	Right	6	Cones blocking	Diverted route
161	B	Right	6	Cones blocking	Diverted route
182	B	Right	6	Cones blocking	Diverted route
187	B	Right	6	Cones blocking	Diverted route
189	B	Right	6	Signs/signals	Signs/signals
190	B	Right	6	Signs/signals	Signs/signals

Paper 6

Effect of curved vs flat screens on remote operator SA for forklift operation: A collaborative study on remote operation.

Abstract

This project is a research collaboration between Phantom Auto and Royal Holloway, University of London (RHUL) to identify the effect of type of screen presentation (flat monitor/curved monitor) on remote operator SA for forklift operation. A within-participants design consisting of three tasks (load, unload and cross dock) measured operator performance time in seconds. We used a linear mixed effects model with Condition and Task as fixed factors and subject and pallet load type as random factors. The model showed a significant main effect of Condition ($\beta = - 0.07$, $t = -4.31$, $p < 0.001$, 95% CI [-0.10, -0.04]). Performance time in the flat screen condition was 93% of the time taken in the curved screen condition, with the trailer unload task being the fastest task to complete and the cross dock task taking the most time. Qualitative findings showed participant variations in driving style and strategies for remote driving, learning effects through observation and exposure to other operators driving strategies each had an impact on the perception, comprehension, and prediction of SA for remote forklift operation. Although manual operation was associated with faster performance times than remote operation, qualitative findings in this study indicate that remote operation confers benefits which may outweigh lower performance speeds such as the potential to augment visual perception through variable video feeds. This study recognises the potential to overcome the differences in SA between in situ and remote operators of forklifts by harnessing advances in digital technology and increasing scale in manufacture to sensitively design interfaces that have been empirically proven to support remote SA.

1.0 Introduction

Remote operation or teleoperation is increasingly becoming a normalised part of digital society. Working remotely presents an opportunity for skilled experts to advise, monitor or work at sites which have previously presented a hazard to human health and safety (Almeida, Menezes, & Dias, 2020, Almeida, Patrao, Menezes, & Dias, 2014). Information about the environment from sensors such as cameras and lidar, placed around remotely operated machines, is presented to the remote operator through a graphical user interface enabling them to control and manoeuvre the machine from a remote location (Linkov & Vanžura, 2021). Through the adoption of teleoperation systems, human operators can communicate and control highly automated vehicles (Daw, Hampshire, &

Pender, 2019), operate agricultural machines (Iinuma et al., 2020), fly drones (Chen, Ulmer, & Thomas, 2022) or remotely pilot forklifts in warehouses and cold storage (Costa, 2021).

The adoption of automated guided vehicles and remote operation can increase accessibility to dangerous environments, such as cold storage, and can facilitate longer operating hours by avoiding the need for workers to spend time within the environments in person. It will open the labour pool, offering the potential for greater productivity through 24-hour global shift work as time zone and location restrictions are lifted. If workers can be employed in multiple warehouses, porting between them with a click of a button, existing labour shortages across the warehouse logistics industry can be resolved (UKHaulier, 2021). Other potential benefits of remote warehouse working include increased safety for employees, lower compensation insurance costs for employers and increased commercial output by having less idle time of material handling equipment.

1.1 Research collaboration between Phantom Auto and Royal Holloway, University of London (RHUL) and TRL

Phantom Auto is a Silicon Valley start-up with three primary verticals: Remotely Operated Forklifts, Assistance of Autonomous Solutions, and Distance Driver Training. Phantom's platform utilises drive-by-wire operation to remotely control electric lift trucks from thousands of miles away (Michel, 2022). Phantom Auto plan to deploy thousands of remote-enabled forklifts over the next several years in partnerships with ArcBest and NFI.[1]

Phantom Auto is currently carrying out tests remotely operating electric pallet jacks in different warehouses from their South San Francisco office. A pallet jack is the most basic form of forklift used to load or unload a trailer or move pallets to another section of a warehouse. It works by using tapered forks that slot beneath pallets, then operators raise or lower the pallets before driving. Data and interviews suggest a manned pallet jack with an experienced operator can load 24 pallets weighing 500lbs (non-stacked) onto a 53ft truck in 30 minutes or less. An inexperienced operator will take closer to 40 minutes. Currently, Phantom’s goal is to achieve similar throughput numbers utilizing a variety of methods including technological advances, availability of vehicles and operators, and training. This research project represents a research collaboration between Phantom Auto and Royal Holloway, University of London (RHUL) to identify factors contributing to operator

---

[1] https://www.wsj.com/articles/freight-operators-plan-to-deploy-thousands-of-remote-operated-forklifts-11642597202

performance, enabling product decisions that are based on empirical data with the aim of increasing remote operators' situational awareness (SA).

1.2 Situation awareness in remote operation of forklifts

There are numerous definitions of SA (see Endsley, Bolte, & Jones, 2003; Gugerty, 1997; Lo, Sehic, Brookhuis, & Meijer, 2016; Niklasson et al., 2007) but the most commonly acknowledged is Endsley's model, which considers people's awareness of their environments in terms of perception, comprehension and projection. More simply, SA can be described as what you can perceive around you, what you understand is currently happening and your predictions of what might change in the near future. Remotely operating automated vehicles has colloquially been likened to 'driving blind' as remote systems are currently unable to precisely match the sensory data available to drivers present in the vehicle (Rosenzweig, 2020). This highlights the importance of developing an understanding of the factors contributing to SA specifically in remote operators.

For example, unlike in-situ operators, remote operators are reliant on video feeds being transmitted to them from the environment with which they are interacting. Network latencies may be subject to occasional lagging, meaning that the video feeds used to build remote SA may not always be perceptually or temporally accurate (Zhou, Wang, Wan, & Qi, 2020). However, variable latency spikes can be more damaging to SA than stable high latency as operators can at least learn to adapt to degraded visual information (Law Commission of England and Wales, 2022). When the transmission of visual feeds falls below a certain acceptable threshold, it forces operators to drive slower, thus affecting performance. Phantom Auto[2] asserts that its patented software "*seamlessly aggregates all available networks including LTE, WiFI, 5G, and more and dynamically adjusts to network fluctuations in real-time to deliver the best remote operator experience*" demonstrating the importance that remote operation companies place on reliable systems connectivity

Another challenge that is faced uniquely by remote operators relates to the fact that building and maintaining remote SA is contingent on the capacity of the teleoperation interface to provide a sense of telepresence (Almeida et al., 2020). Presence and task performance have been found to be positively correlated in studies examining behaviour in virtual environments (Schubert, Friedmann, & Regenbrecht, 2001). Qualitative measures to inspect feelings of presence in remote environments have focused on the subjective experience of "being there" and whether the remote environment felt 'real' to judge whether an interface creates a sense of telepresence in the user (Schubert et al.,

---

[2] Phantom Auto - Remote Operation for Logistics

2001). Ideally, a remote operator should feel as if they are inside the vehicle, picking up the same sensory and environmental cues that they would if they were physically present at the machine's location (Almeida, Patrao, Menezes, & Dias, 2014). This may still be some way in the future until technology can faithfully recreate that experience. For now, the best determinant of a remote operator's sense of presence might be that they felt they were where the machine is, rather than sitting at their control station (Witmer & Singer, 1998). It is likely that this sense of presence can be increased by exploring alternative approaches to presenting the remote operating interface at the control station.

1.3 Comparing effect of flat or curved monitor on remote operator performance times

Phantom Auto's remote operators currently use a flat screen 32” monitor to view the visual feedback from the pallet jack via a browser window which hosts their proprietary operating software. Phantom Auto recently acquired Voysys, a video communication solution for teleoperation over public networks, and is in the process of combining the symbiotic technologies to enhance performance. Voysys asserts that they have achieved an increase in productivity of up to 25% by switching to a larger, curved monitor which their operators claim provides more of a sense of presence in the remote environment and provides better situation awareness cues.

Providing operators with a wider viewing display has been explored in many different industries for many years as a method for increasing situation awareness, for example giving pilots panoramic screens to enhance visual perception during flight operations (Möller, Kostka, Neujahr, & Engineering, 2002), video display walls in control centres (Snavely & Patterson, 2008) and for sim car racing in gaming industries using curved screens to widen the field of view (FOV) for enhanced driving performance (Finn, 2021).

In the human visual system, our eyes have a horizontal FOV (hFOV) of around 190°, including the peripheral information we can perceive on either side of our vision when we are looking straight ahead, and a vertical FOV (vFOV) of about 135° (Lidestam, Eriksson, & Eriksson, 2019). Flat screen monitors mainly display visual information directly in front of the viewer, whereas, in curved monitors, the edges of the screen bend around them, providing a wider viewing distance across the screen and access to more peripheral information from the camera view (Kyung, 2019) .

Manipulating the FOV can influence how we perceive the optic flow in our environment, altering our behaviour within it. Increasing FOV can modify the way in which people drive; by degrading accurate speed perception people feel they are driving too fast and will slow down (Lidestam et al., 2019). The optic flow in the real world is difficult to match with computer monitors, so we are dependent on the FOV that is presented to us via the display monitor to gather information from the remote scene to determine our actions.

Providing a wider FOV in a monitor display has been shown to impact positively on performance in visual search tasks. Kyung (2019) measured the effect of different display curvature radius between 400R, 600R, 1200R, and flat (where 400R represents a display curvature radius of 40cm) and display size (33" and 50") on a visual search task. They found that larger display settings on flat screens had a detrimental effect on performance and the most accurate performance was demonstrated at 600R for curved screens for both display sizes, concluding that wide flat screens are not advisable in visual display dependent tasks. However, Klatt & Smeeton (2020) found no differences in perception or decision making when stimuli were presented on either flat or curved screens. However, they concluded that curved screens were better for attention-based tasks where viewing quality was critical, as the shape of the display resulted in less image distortion than flat screens.

Phantom Auto and RHUL aimed to empirically determine whether remotely operating a pallet jack via a 49" curved screen (hFOV 95°, vFOV 63°) or a 32" flat screen (hFOV 71°, vFOV 43°) has an effect on the time it takes for operators to carry out three tasks frequently executed in standard warehouse operations: a load task where ROs load pallets into a trailer, an unload task in which they unload pallets to another area in the warehouse, and a cross dock task which requires ROs to move pallets between aisles of the warehouse.

2.0 Method

2.1 Participants

The sample was originally planned to consist of eight operators, but logistical issues meant that only five operators were tested. All were in current employment at the company, working in the South San Francisco office. They had a mean age of 37.4 years. Each operator had different levels of experience in operating forklifts both manually and remotely. Three also had manual experience on the pallet jack.

Table 1 Demographic data for n= 5 participants

	Overall (N = 5)
Mean age (in years)	37.4
Gender	
Female	0 (0%)
Male	5 (100%)
Mean no of years forklift experience	3.67
Mean no of years remote driving experience	2.2

Participants were recruited on a voluntary basis and were under no contractual obligation to take part in the study. They were tested during normal office hours and other colleagues took over their work tasks for the period of the study. All procedures were reviewed and approved by the Royal Holloway Research Ethics Committee.

2.2 Materials

2.2.1 Hardware

The experimental set-up and hardware were identical in each condition. Both operator workstations were set up inside the offices which were housed directly behind the warehouse wall. It was not possible to hear sounds from inside the warehouse although there were other employees further down the office working at their desks.

In the curved screen condition, we used a 15.08" x 47.36" x 20.68" SAMSUNG 49-Inch CRG9 Curved Gaming Monitor with 120Hz refresh rate and dual QHD 5120 x 1440 resolution with a screen curvature of 1800R. For the flat screen condition, we used a 29.2" x 17.4 "x 2.7" SAMSUNG 32-Inch SD850 WQHD Monitor with 60Hz refresh rate and QHD 2560 x 1440 resolution.

We used a Unicarrier SPXE80 35.1" x 56.9" chassis pallet jack for remote operation which has a basic capacity of 8,000lbs and a maximum forklift height of 9.25 inches fitted with a V21Intel 10th Gen Quad, Core i7/i5/i3 Fanless Embedded System onboard computer. It had seven on board cameras; left, right, front, and rear mast cameras, right and left side cameras and the fork camera (see Figure 1 and Table 2 for details).

REDACTED FOR COMMERICAL PRIVACY

Figure 1 Details of Unicarrier SPXE80 and camera positions

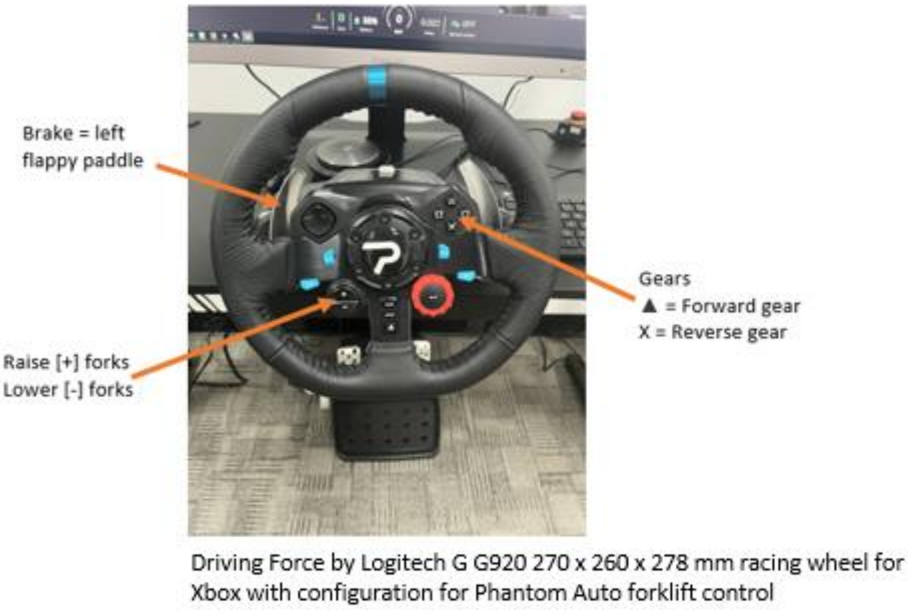
Table 2 Camera orientation, pitch and resolution details on pallet jack

REDACTED FOR COMMERICAL PRIVACY

OBS Studio[3], a free and open source software package for video recording and live streaming recording was used to capture high performance real time video/audio of each drive. Pallet pick up/put down was recorded on an Accusplit Pro Survivor 601X stopwatch in min:sec:1/100s and converted to seconds for analysis.

2.2.2 Lab set-up

The operator sat in front of the steering wheel which was secured equidistant to each side of the desk (52.07cm lateral and 52.07cm longitudinal) and 63.5cm from back of the desk. The steering mechanism was a commercially available gaming racing wheel for Xbox (Driving Force by Logitech G G920 270 x 260 x 278 mm) which had been configured to Windows 10 to control the forklift fork raise/lower mechanism and change gears from forward driving to rear driving (fork side). See Figure 2 for details.



Driving Force by Logitech G G920 270 x 260 x 278 mm racing wheel for Xbox with configuration for Phantom Auto forklift control

Figure 2 Steering mechanism for remote operation of pallet jack

---

[3] https://obsproject.com/

a) key board and mouse with audio microphone b) foot pedal (clutch, accelerator, brake (left to right) c) remote operator work station curved (left) and flat screen (right)
n.b. note relative height of monitors and position of steering wheel

Figure 3 Keyboard, pedals and monitor layout

The pedal plate (H167xW 428.5xD311mm) was a standard vehicle pedal control. The accelerator pedal was used to control throttle speed and the brake pedal to stop the forklift. The clutch pedal was non-operational as the gears were changed using a button on the steering wheel. The foot pedals were placed directly underneath the desk in line with the steering wheel and pushed flush against the back wall, so the pedal plate did not move during testing.

Each monitor was placed 10" behind the centre of the steering wheel. A keyboard and mouse situated to the left of the steering mechanism was used to start the screen recordings together with the audio speaker and microphone which was used to communicate with the safety monitor in the warehouse when the audio channel was open. See Figure 3 for details.

The chair position relative to the desk and seat height was adjusted to operator driving comfort based on operator height and leg position to reach the pedals. A chest bar was used to measure 66.04cm distance from the middle of the screen to the operator's chest and the chair position was marked on the floor to ensure that the operator did not change position during testing or after breaks. The same chair was used in both conditions.

2.2.3 Warehouse set up

The warehouse was fully operational with a section marked out for testing which was not used by others for the duration of the testing period. See Figure 4 for diagram of the warehouse layout.

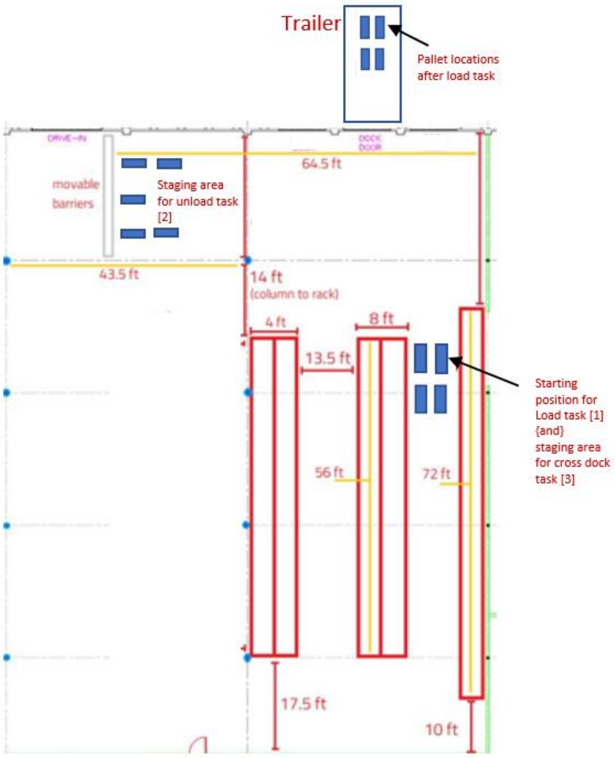


Figure 4 Warehouse layout and location of pallets for each task nb. measurements are imperial due to US location.

The testing site consisted of

a) a 161.54m trailer housed in front of an open bay door used for the load task [1] and unload task [2],

b) a 8.89m x 5.11m staging area used to place pallets unloaded from the trailer with three columns marked on the floor in yellow tape used for the unload task [2],

c) two 24.38m aisles divided by storage racks

d) a 16.12m x 2.59m staging area on the far aisle divided into two columns marked on the floor in yellow tape used for the 'cross dock' task [3]. This was also the starting position for pallets for the load task [1] (see Figure 5 for visual details of testing site).

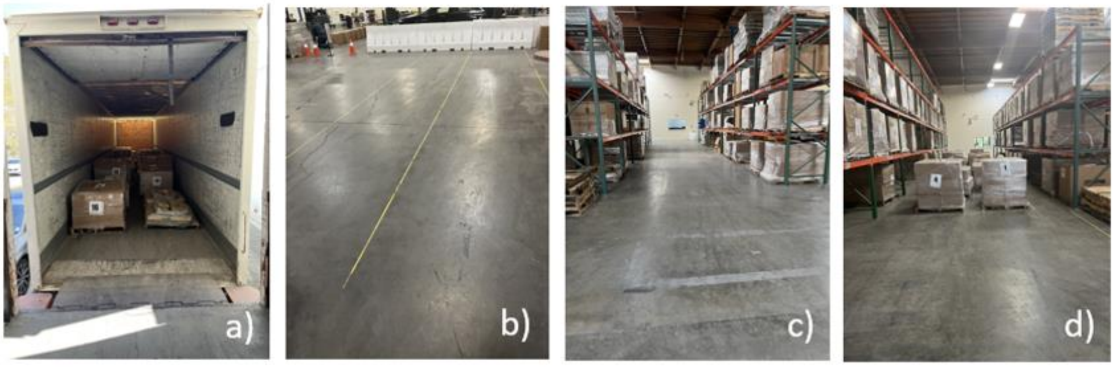


Figure 5 Visual representation of each of the three tasks

2.2.4 Pallet loads

A typical trailer load delivery of pallets to a warehouse may have multiple pallet loads of different types of specification. This variability in load type was maintained in the testing set-up to ensure mundane realism. All 22 pallets were standard 101.6 x 119.38cm warehouse pallets with approx. 226.80kg loads constructed from high or low stacked boxes wrapped in cellophane, cement bags secured with straps and cinder blocks or combinations of loads (see Figure 6 for examples of each pallet type). Pallets were numbered and their load type recorded (see Table 3). If a pallet was broken during testing it was replaced by the same pallet load type.

Table 3 Pallet load type and numbered pallets of that type

Pallet load type	Pallet numbers
High boxes (106.68cm from bottom of pallet)	1, 4, 7, 18, 19, 21
Low boxes (74.93cm from bottom of pallet)	2, 3, 5, 6, 10, 11, 15, 16, 20
Cement bags and cinder blocks	8
Cinder blocks	9, 12, 22
Cement bags	13, 14, 17

a) Standard pallet b) low boxes c) cement bags and cinder blocks d) high boxes e) cinder blocks f) cement bags

Figure 6 Visual examples of each pallet type

2.2.5 Interface layout

The interface for remote forklift driving can display seven cameras; left, right, front, and rear mast cameras, right and left side cameras and the fork camera. The location of the cameras, their number, and layout on the operator's screen is usually configured to the operator's preferred layout. However, for this study we agreed on a fixed layout through an informal focus group with four experienced remote forklift drivers to establish a layout for both screens. There were limitations on re-sizing camera images to fit both monitors; a small aperture on the flat screen could mean drivers would struggle to perceive fine details in the remote scene but stretching the image to fit the curved monitor could result in the image having low resolution and being too blurry to drive safely. The layout for the curved screen condition can be seen in Figure 7 and for the Flat screen condition in Figure 8.

In the curved screen condition layout, to use the full curvature of the curved monitor, the left/right side cameras were placed in the curved edge sections and the size of the left, right, and forward mast cameras were stretched to fill the space but retain resolution. In the flat screen condition layout, if the configuration was the same layout as the curved monitor, the images were too small to operate properly. Instead, the left, right and front mast cameras were stretched across the top of the screen and the left/right side cameras were placed directly underneath. As a result, the interface layouts were as similar as was practically possible while facilitating safe driving conditions but were not fully matched between conditions.

In both layouts, the main on-screen centre camera presented the front/rear mast camera, which switched positions when the operator changed gears showing the forklift direction of travel. The fork camera was placed underneath the front and right mast cameras, as operators reported that they only used this camera when seating the load on the pallet, not during active driving. Extra details in the interface design showed operators the current gear and the forklift driving speed. The layout for both conditions was held constant across participants.

Figure 7 Curved screen condition interface layout

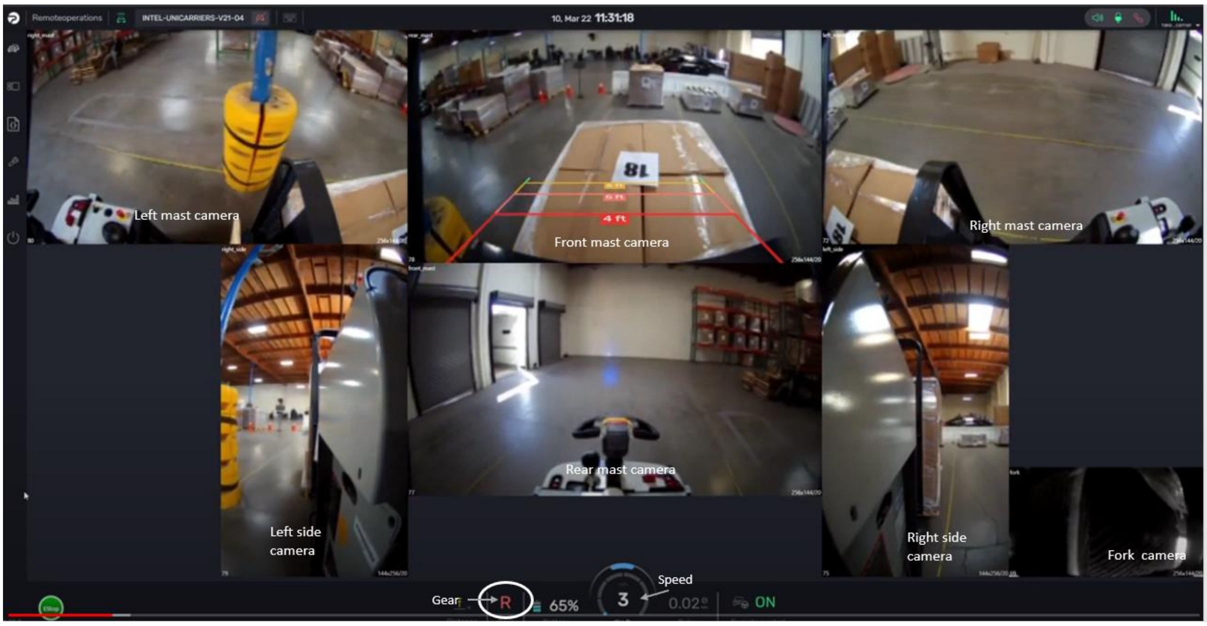


Figure 8 Flat screen condition interface layout

2.3 Design

A within-participants design was used to test the effect of type of screen presentation (flat monitor/curved monitor). Each condition consisted of three tasks in the order, load [1], unload [2] cross dock [3] with condition order counterbalanced across operators. It was not logistically possible to rotate the order of the 3 tasks in the conditions as that would mean manually restacking all pallets into a different configuration, which took over an hour at a time to reorder.

2.4 Dependent measures

The dependent variable was operator performance measured by time in seconds. Performance was operationalised as the time from pallet pick up to pallet put down in seconds. This shows how much time operators interact with each pallet and allows the effect of pallet load to be assessed.
The experimenter observed operators driving via the remote interface and timed the pallet pick up/put down on a stopwatch.

Further exploratory analysis was carried out after the testing had taken place comparing the performance time for each participant in the flat screen condition for the load task under remote conditions (teleoperation via GUI) with the time taken for them to conduct the load task under manual conditions (on-board operation).

2.5 Procedure

Participants completed ethics and consent and answered demographic questions hosted on Gorilla.sc about their remote driving and forklift experience. They were informed which order they would be completing the experimental conditions and seated at either the flat 32" monitor or the curved 48" monitor station. Participants adjusted the seat height and chair position relative to the pedals to a comfortable driving position. A chest bar was used to check that the participant was sat approximately 26" from the monitor. As a safety precaution, before each condition started, they were given a ten-minute practice session in a corner of the warehouse driving the forklift backwards and forwards and picking up/putting down a pallet to familiarise themselves with the layout of the cameras in that condition.

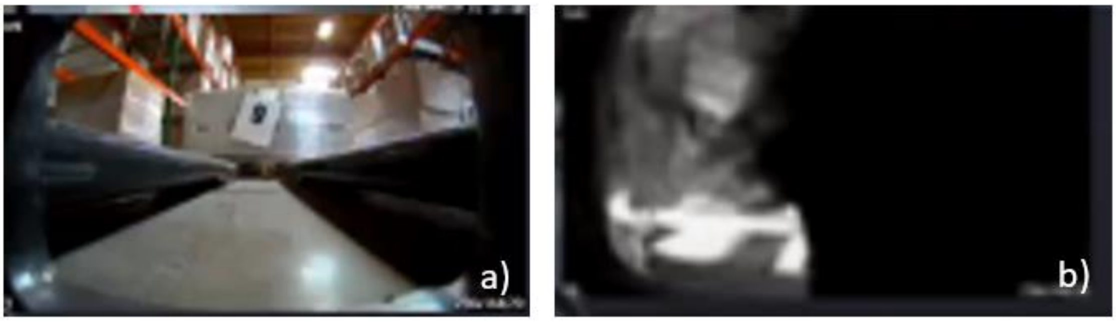
Participants carried out three tasks in each condition in the same order: load [1], unload [2] and cross dock [3]. Each task required a set of standard manouvres to carry out the task (see Table 4).

Table 4 Task description and driver manoeuvres required to carry out the task

Task	Manouvres required to carry out the task Nb. When forks are forward (forkside), operators are driving in reverse gear.
Load [1] - drivers were required to move 22 pallets individually from a staging area marked out on the floor in the warehouse to inside a 53' trailer located XXXX away.	• Activate the reverse gear by pressing X on the steering wheel then drive towards the pallet with lowered forks. • Insert the forks fully into the gaps under the pallet until the load completely obscures the fork camera at the back of the pallet forks, activate the forks using the + button on the steering wheel to lift the pallet. • Change to forward gear using the ▲ button on the steering wheel and drive towards the bay door (pulling the pallet). • Complete a two-point turn to reorientate the forklift into driving forks forward, change gear to reverse and enter the trailer negotiating the side of the bay door safely. • Once inside, operators precisely placed each pallet into position starting at the back into 2 columns negotiating the trailer walls and placing the pallet close enough to the other pallets or back wall to ensure that 22 pallets would fit inside the trailer. • Once positioned, lower the forks using the – button on the steering wheel, engage the forward gear and drive forward to extract forks from pallet. • Exit the trailer and drive back to the staging area to collect the next pallet.
Unload [2] – drivers were required to unload 22 pallets from a 53' trailer to a staging area marked out on the floor in the warehouse.	• Activate the reverse gear by pressing X on the steering wheel then drive into the trailer towards the pallet with lowered forks. • Insert the forks fully into the gaps under the pallet until the load completely obscures the fork camera at the back of the pallet forks, activate the forks using the + button on the steering wheel to lift the pallet. • Change to forward gear using the ▲ button on the steering wheel and drive out of the trailer (pulling the pallet).

	• Complete a two-point turn to reorientate the forklift into driving forks forward, change gear to reverse and drive towards the 3-column staging area in the warehouse • Place the pallet in one column negotiating other pallets and/or the wrapped warehouse column strut. • Once positioned, lower the forks using the – button on the steering wheel, engage the forward gear and drive forward to extract forks from pallet. • Drive back to the trailer to collect the next pallet.
Cross dock [3] – drivers were required to move 22 pallets from a staging area marked out on the floor of the warehouse to another staging area located in one of the aisles of the warehouse.	• Activate the reverse gear by pressing X on the steering wheel then drive towards the pallet with lowered forks. • Insert the forks fully into the gaps under the pallet until the load completely obscures the fork camera at the back of the pallet forks, activate the forks using the + button on the steering wheel to lift the pallet. • Change to forward gear using the ▲ button on the steering wheel and drive away from the staging area. • In front of the first aisle, complete a two-point turn to direct the forklift into the first aisle. • Drive down the first aisle until the end of the racks and turn left. Drive into the far corner of the warehouse, stopping at the wall. • Change into reverse gear and drive reverse, lining up with the second aisle. • Drive forkside towards the 2-column staging area at the start of the second aisle. • Place the pallet in one side of the 2-columns (behind other pallets if present). • Once positioned, lower the forks using the – button on the steering wheel, engage the forward gear and drive forward to extract forks from pallet • Drive back to the staging area to collect the next pallet.

The time to move each pallet from its starting position to its final position was recorded as performance time. The experimenter sat behind the participant on the right-hand side and recorded each pallet pick up by pressing start on the stopwatch when they observed the participants pressing the raise forks button [+] and saw that the fork camera was fully obscured (see Figure 9 for images of the fork camera with and without load). When the participant lowered the forks by pressing the lower forks button [-] and the experimenter saw the forks pulling out of the pallet, they pressed stop.



a) Fork camera with no load b) fork camera with fully seated load (occludes the camera)

Figure 9 Fork camera with no load and fully seated load

The starting position of each numbered pallet and its placement in the trailer or staging area was recorded. On occasions when there were stoppages mid-move, due to broken pallets or other factors, the stopwatch was paused, and the participant informed that recording has been halted. All unintended disruptions such as pallet breakages were recorded. When the operator resumed moving the forklift the timer was restarted. Live streams of all drives were recorded by OBS, and the recorded pallet times and total stoppage time were verified by checking video recordings after the drive.

After each task was finished, participants were allowed ten-minutes to step away from the monitor and/or take a comfort break. All three tasks for the condition were completed in one session. At the end of each condition, participants completed the NASA TLX-R Workload measures (Hart, 2006) indicating answers on a sliding scale where responses were quantified from 'very low' (0) to 'very high' (100). To measure their sense of presence in that condition, two questions were adapted from Schubert, Friedmann, & Regenbrecht's (2001) presence scale. The question, "To what extent do you agree with the statement, "*I had a sense of being in the warehouse while carrying out the task/s in this condition*" was measured on a Likert scale from "strongly agree" (0) to "strongly disagree" (5). The question " *While you were operating the forklift using the images displayed on a curved screen* [or flat screen depending on condition], *how real did the warehouse feel to you*?" was answered on a sliding scale from 'very low' (0) to 'very high' (100). Five questions measuring cyber sickness symptoms were adapted from the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, 1993) with each item rated on a scale of from none [0], slight [1], moderate [2], to severe [3].

The following day, operators returned to the office and completed the same procedure for the second condition. See Figure 10 for a diagram of the procedure.

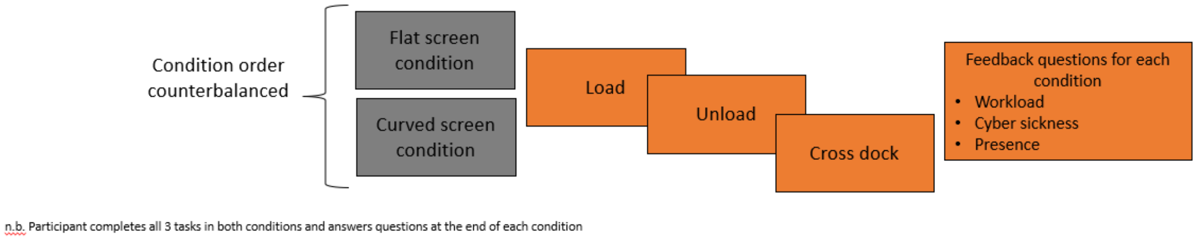


Figure 10 Procedure of the experiment n.b. Participant completes all 3 tasks in both conditions and answers questions at the end of each condition

3.0 Results

3.1 Quantitative results

Descriptive statistics showed that there were shorter times overall for the unload task and the cross-dock task took longest in both conditions. There were variations between operator performance, although variation between different drivers appears to be lower with the curved screen which could be an effect of completing that condition first (see Figure 11). Due to time and recruitment limitations, 3 operators carried out the curved screen condition first, followed by the flat screen condition, and 2 operators conducted flat, then curved conditions meaning the dataset was not fully counter balanced.

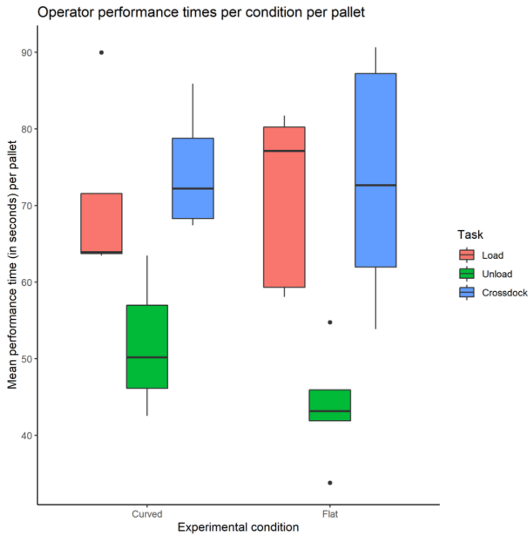


Figure 11 Comparison of operator mean performance per condition

3.1.1 Analysis 1 - time from pallet pick up to pallet put down in seconds.

We used a linear mixed-effects model (LMM) as it can naturally handle unbalanced datasets with repeated measures. Condition (Flat screen /Curved screen) and the Task (Load/Unload/Crossdock) were entered as a fixed effects. We accounted for individual differences between drivers and for variations in pallet load type by adding these factors as random intercepts in the model[4]. The dependent variable was operator performance time, measured by time from pallet pick up to pallet put down in seconds. P-values were estimated using Satterthwaite's approximation as implemented in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

---

[4] lmer(rt_log ~ Condition + Task + (1|subj) + (1|Pallet_type))

Visual inspection of the model fit showed that the residuals were positively skewed, so we applied a log transformation to the performance time variable. We used parametric bootstrapping to estimate confidence intervals and evaluate statistical significance.

There was a significant main effect of Condition (β = - 0.07, t = -4.31, p <0.001, 95% CI [-0.10, -0.04]). Exponentiated coefficient values can be interpreted as predicted multiplicative changes in the dependent variable across conditions (since the dependent variable was log-transformed). Using this approach, the model indicates that the flat screen condition was associated with a performance time that was 93% of the time taken time in the curved screen condition. There was also a significant effect of TaskUnload (β = - 0.37, t = -18.81, p <0.001, 95% CI [-0.41, -0.33]) and TaskCrossdock (β = 0.07, t = 3.36, p <0.001, 95% CI [0.03, 0.11]). On average, drivers unloaded the trailer in 63% of the time it took to load it and were slightly slower at completing the cross dock task than at loading the trailer (about 7% increase in performance time). The cross dock time is predominantly impacted by the extra distance travelled.

Random coefficients of the model showed that the high box pallet type took 4.25% longer to remotely operate than the other pallet types, which is likely due to the extra 12.5" height of the pallet load preventing remote operators from being able to see the position of the other pallets in relation to it.

3.1.2 Analysis of Presence measures

Participants were asked two questions after each condition to measure their sense of presence in the remote scenes. There was no significant difference (t(5) = 0.54, p = 0.60) between the perceived 'realness' of the virtual world in the flat screen condition (M^{Flat} = 22.6) and the curved condition (M^{Curved}= 31.2) for the scaled question " *While you were operating the forklift using the images displayed on a curved screen* [or flat screen]*, how real did the warehouse feel to you*?" (see Figure 12).

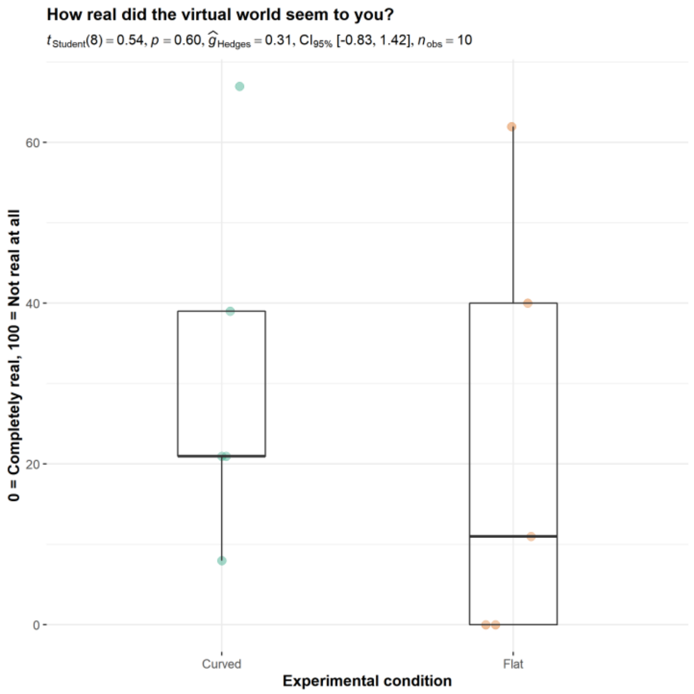


Figure 12 Participant responses to the Presence question across condition

There was no significant difference (Fisher's exact test p = 1) in responses to the question, "To what extent do you agree with the statement, "*I had a sense of being in the warehouse while carrying out the task/s in this condition*".

3.1.3 Analysis of Cyber sickness measures

There was no significant difference (t<1) between any of the self-reported cyber symptoms (eye strain, headache, fatigue, dizziness or headache) between the two conditions. All drivers responded either "None" or "Slight", irrespective of condition.

3.1.4 Analysis of Workload measures

There were no significant differences between the two conditions for any of the responses to the NASA-TLX questions (t<1). Drivers reported that high levels of attention and visual demand were required in both conditions and lower levels of physical and temporal demand and stress, irrespective of whether they were using a flat or curved monitor (see Figure 13).

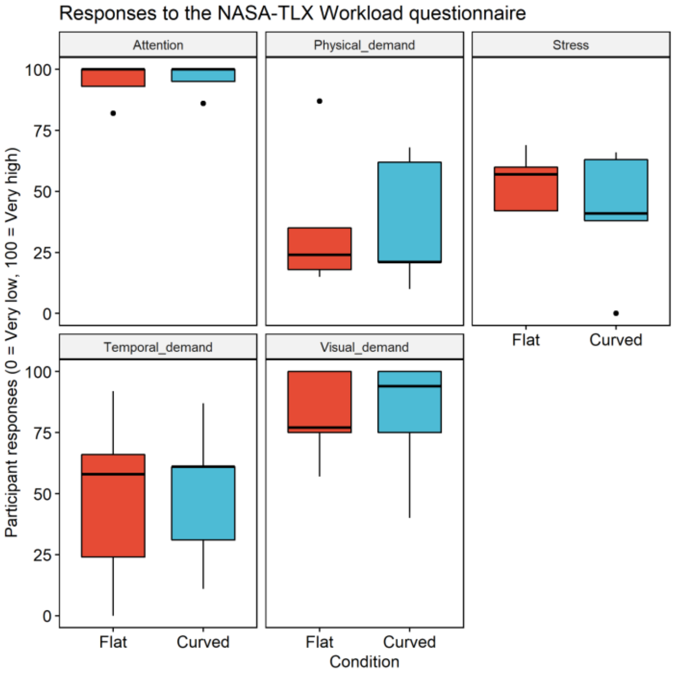


Figure 13 Participant responses to NASA-TLX questions across condition

3.1.5 Exploratory analysis of remote operation compared to manual operation of the load task

Participants carried out the load task under manual operation conditions (on-board the pallet jack) and we compared their performance times to their data for the Flat screen condition load task in Analysis 1. We conducted the same LMM model with only Condition (Flat screen /Curved screen) as a fixed factor and pallet type and subject as random factors[5] and applied the log transformation to the performance time variable again.

---

[5] lmer(rt_log ~ Condition + (1|subj) + (1|Pallet_type))

We found a significant main effect of Condition ($\beta = -.64$, $t = -12.98$, $p < 0.001$, 95% CI [-0.75, -0.55]). The model indicates that operation under manual conditions was associated with a performance time for the load task that was 53% of the time taken time under remote conditions. This is however a prototype vehicle and Phantom has a roadmap to close that gap through technology improvements, operational solutions, and training.

3.2 Qualitative results

In addition to measuring driver performance quantitatively, we took qualitative observations across a wide breadth of topics during the testing period. The following section focuses on the overarching patterns uncovered which demonstrate the range of challenges involved for operators to build and maintain enough SA to operate a pallet jack remotely.

3.2.1. Participant variations in driving style and strategies for remote driving

At the time of writing, there is no statutory or regulatory guidance in the UK or US that directly mandates how remote operators of pallet jacks should perform driving tasks. Remote operation can be variable, personalised and is often dependent on operators' previous training. Participants in this study demonstrated numerous strategies for performing the same task, for example how operators 'seat the load' was highly individualistic. To move a pallet, the pallet forks must be fully inserted inside the pallet and raised to their full height to avoid dragging it along the warehouse floor. Operators adopted diverse techniques to complete this exercise; one example is to activate the forks to lift the pallet, but within the same manoeuvre, immediately start moving backwards so the pallet is lifted to its height whilst in motion, effectively dragging it up and out of position in one seamless move. This was justified by those who used it as the soundest strategy, ensuring that the load was fully seated, enabling a faster pallet pick-up and limiting pallet breakage. Another, opposing strategy, is to carefully insert the forks and wait until the forks were lifted to the full height before beginning to move backwards. Avoiding pallet breakage was also given as the rationale for this action. Whichever strategy was adopted, operators had full confidence that their technique was advisable protocol for other drivers.

Further individualised behaviours were observed in driving styles to complete specific manouvres. To complete the load task, some operators performed a careful 2-point turn towards the bay door, stopping to use the forward and reverse gears to move into place, whereas others used a swinging motion whilst driving backwards, dynamically pulling the pallet jack into position before activating the forward gear. Participants also drove at differing speeds (particularly when driving down the first aisle in the cross-dock task) with a range of 3-9 km/h between operators. This behaviour was observed to be consistent *within* operators in both conditions. These inconsistencies between drivers did not appear to significantly impact on performance, yet future training protocols would profit from identifying the most judicious strategies and adopting them as a universal method across operators.

The cross dock task also produced variations in driving technique as to how operators turned the corner between aisles. Most drivers positioned themselves as far into the corner of the warehouse wall as possible, pivoted the forks by changing gears and then drove forkside out of the corner, down the aisle to place the pallet. However, one participant drove forkside down the whole first aisle, allowing them to enter the second aisle faster, without the need for a turning manoeuvre. Usually, there is a speed cap when driving in reverse for safety reasons, but the safety system was turned off for the testing period. Any performance time gains the participant maximised in this study would not translate in an operational warehouse with safety regulations.

3.2.2 Differences between remote and manual operation of pallet jack

We conducted a qualitative comparative assessment of the differences in driving style and operation for manual and remote driving in the load task independent of the flat/curved screen paradigm. There were several indications that there are important variations in SA and driving behaviours between manual and remote operation.

Standing on the pallet jack enables the manual operator to use their height to look down over the next pallet to judge how close they are to it when placing the pallet, enabling quick SA judgements and action, regardless of pallet load. Conversely, when remote driving, load type can entail time consuming analysis for the operator. If the load consists of stacked high boxes, this prevents the operator from seeing how close they are to the next pallet, particularly if the previous pallet load was low, for example cement bags. Consequently, the remote operator must carefully consult all the cameras one-by-one to pick up the necessary cues to make the judgement of the required space between the pallets before moving which impacts on performance times.

The dichotomy between remote and manual operation is also evident when picking up pallets and transporting them. The remote operator must carefully insert the forks into the pallet, check the fork camera to ensure that the load is seated, change gears, which in turn switches the cameras, then use the buttons on the steering wheel to lift the pallet and move back using the mast camera. When manually operating, the pallet is 'scooped' up and moved backwards in one fluid move. Operators simultaneously move the forks up and move backwards, looking over their shoulder. There is no need for a separate check underneath the pallet to ensure the load is seated as the manual operator can observe that the load is flush against the backrest.

There are differences in visual acuity between the manual and remote operation of the pallet jack when dealing with the issue of broken pallets or debris on the warehouse floor. Manual operators can spot a tiny shard on the floor which is difficult to detect remotely even with high resolution cameras and just kick the splinters out of the way themselves in seconds. The benefit of being physically in the warehouse means that they can deal with damaged pallets themselves just by jumping off the pallet jack and checking under the pallet to gauge the integrity of the pallet.

Yet there are areas where remote operation profits from augmentation to the visual feed in comparison to manual operation. When the pallet jack is operating without the safety system enabled, Phantom requires a safety monitor with an emergency stop switch to intervene if there are safety conflicts. The safety system, if it is in use inside the trailer activates if operators get too close to the side wall, which shuts off the pallet jack until the safety monitor manually moves it away from conflict. To avoid this, when operating remotely inside the trailer, the driver must carefully check the side wall using the side cameras, moving cautiously, and ensuring there is a visible gap between the pallet and the trailer wall. During manual driving, operators are standing on the pallet jack and can lean over the tiller to judge distance whilst going forward or look over their shoulder as they reverse. This allowed participants able to move the pallet jack within millimetres of the side wall relatively quickly.

When operating remotely, operators can gain the necessary SA by shifting visual attention from the right and left side cameras to the front three mast cameras, interpreting perceptual cues from the fixed camera angles on the pallet jack. This provides a broader range of visual information but processing and interpreting the data can slow down performance. However, adjustable visual guidelines which can be configured virtually on the operator interface can also be used to indicate the relative depth cues between the end of the forks and the pallet in front of them to support SA during pallet pick up and placement. Virtual enhancements, such as these, has the potential to optimise remote operation performance in comparison to the manual operation, by reducing the guesswork and individual skill required to carry out the load/unload task.

There are also physical differences between manual and remote operation which may impact on the driving manoeuvres that are adopted in each context, even for the same operator. For example, during the cross dock procedure, one participant who drove down the first aisle forkside with no turning manoeuvre at the far corner of the warehouse in the remote operation condition, whereas when operating manually they swung round and changed gears going into the second aisle forkside. When asked why they had changed their behaviour, the operator explained that he preferred not to drive down the first aisle facing backwards as having to look over his shoulder was physically uncomfortable. By contrast, when operating remotely, changing gears automatically switches the camera view, allowing driving forkside as if you were driving headlong. This suggests that the variable visual feed that is available to a remote operator can provide benefits in some tasks.

Differences between techniques employed to load pallets were observed. One technique for quick stacking of pallets and loading/unloading involved lowering the pallet to the ground and then using the driving force of the pallet jack to push forward into another pallet and 'shunt' both further along the ground to ensure that they were packed as closely as possible to each other. Participants with manual pallet jack experience were observed to practise this behaviour regularly whilst operating remotely, whereas those participants with remote operation training did not demonstrate it once. Previous experience in live contexts may embolden remote strategies, producing differences in how operators perform elements of the task, suggesting that recruiting exclusively from labour pools with manual forklift experience to train as remote operators, could be a successful strategy. Potentially those with previous forklift experience could be leveraged to act as instructors, widening the labour pool to those with relevant remote experience in other domains.

Finally, there were differences in how participants stood on the pallet jack when operating manually, some participants faced forkside and looked over their shoulder, when necessary, whereas others rode the pallet side-on with hands and feet placed on either side of the tiller like *"steering a boat*" as this enabled them to use the "*counter thrust to make turns like tacking*"(Participant 4). All the operators expressed the physical discomfort of standing on the pallet jack, particularly when driving over the metal dock plate into the trailer which resulted in violent rocking of the pallet jack. Furthermore, their experience of manual driving was cited by some participants as the reason they were driving slowly over the docking plate when operating remotely, as the noise it made in the audio feed reminded them of the physical effect on the pallet jack in the real world.

> *'Unconsciously or consciously, I am remembering the effect it had on me when I was there'* Participant 2

The importance of the audio channel cannot be underestimated while operating remotely. Remote operators are dependent on audio feeds which place them virtually in the environment. With audio, the operator can hear what the machine is encountering such as a piece of broken pallet stuck beneath the wheel, or the main drive wheel losing traction and spinning on a dock plate. They must then communicate with the warehouse safety monitor to remove it, which increases their stoppage time.

The cumulative effect of these qualitative differences between manual and remote operation impacted significantly on performance times, with lower load time observed across all participants for manual operation. However, evidence has shown that human factors such as physical fatigue have a strong impact on warehouse system performance, so the gains in the physical comfort and safety of operators in remote contexts, not to mention the host of other potential advantages outlined above, justify a continued exploration of remote warehouse logistics (Choi, Ahn, & Seo, 2020).

3.2.3 Learning effects through observation and exposure to other operators driving strategies

Social learning theory outlines the importance that environment and cognition have on learning and behaviour and is a key framework for understanding how employees learn to perform the skills required in their workplace as by observing behaviours with desirable outcomes, and imitating them, the observer can minimise damaging and avoidable errors (Fryling, Johnston, & Hayes, 2011). In this study, participants who were regularly in the warehouse, either to carry out other tasks when they weren't part of the study or to perform the safety monitor role, witnessed a range of driving mistakes and successes by other drivers. Adopting the modelled successful behaviour improved their performance times when they were remotely operating later. For example, one participant openly admitted they were "*trying something new*" after seeing the previous driver back the pallet jack into the far warehouse wall during the cross dock task which positively impacted their performance time for that task. Another operator, after observing the difficulty that other drivers had had getting the last two pallets off the forks in the trailer, performed the task first time without any necessary adjustments despite the challenging manoeuvre but commented that they *'had been thinking all morning about how they were going to do that'*. This demonstrates that learning effects through observation and exposure can help to reduce the errors that drivers make when operating remotely.

3.2.4 Situation awareness needs of remote operators of pallet jacks

Occasional latency spikes, audio feedback interruptions and the presence of pedestrians in the warehouse created a series of 'SA demons' for the participants, together with attention tunnelling in the curved screen condition (Endsley et al., 2003).

Latency in the video feed results in a delay between the operator's dynamic driving actions and visual feedback of those actions on the screen, making it difficult for teleoperators to drive smoothly (Linkov & Vanžura, 2021). Accompanying cameras can provide them with enough information to maintain SA and continue driving, but only if operators are confident that the video feed will return quickly. When this occurs in the primary camera, adept operators will recognize the condition and use other available cameras to maintain SA with minimal impact. When latency occurs in more than one camera, this can become increasingly difficult to overcome and operators will slow down until the video feeds catch up.

> "*There was a lot of camera latency all throughout the exercise which forced me to be visually agile while in motion to find a functioning camera. I found myself pretty quickly able to adapt in spite of this*" Flat condition, Participant 3.

Participants overwhelmingly preferred the flat screen configuration which was the usual layout used for remote operation at Phantom Auto, as it provided all the camera images directly behind the steering controls, so the operator only had to move their eyes to access information:

> *"I felt that the camera placement of the mast-front and mast-rear being the same size and orientation helped me make turns well and calculate my next move when switching gears."* Flat condition, Participant 1

Participants commented that the larger size of the front/rear mast camera position in the curved screen condition *'sucked you in'* which had a negative effect on attention and meant participants changed their normal operating behaviour:

> "*Curved leaves basically the centre camera as your main/exclusive view. I found I used less cameras while using the curved screen than while I was using the flat one.* " Curved condition, Participant 4.
>
> *"I felt that I was only paying attention to one camera at a time, and I was not moving my focus to other cameras as much."* Curved condition. Participant 1

One consequence of this attention tunnelling meant that, during a latency spike of 3-4 seconds in the front mast screen during the curved condition, the participant did not deal with crucial information from the other cameras which were showing that he was driving into the warehouse racks, instead using the delayed information as if it were a live feed. The driver was visibly confused and grew increasingly alarmed as he could tell from the force feedback that the pallet jack was no longer responding to his driving control despite the front mast camera showing him driving down the centre of the aisle. Such latency incidents inevitably resulted in over cautious driving afterwards as the participant no longer trusted the visual information that they were receiving. This demonstrates that it is vital teleoperators can trust interface controls and displays to provide accurate information as it clearly negatively impacts on performance, productivity and workload stress (Linkov & Vanžura, 2021).

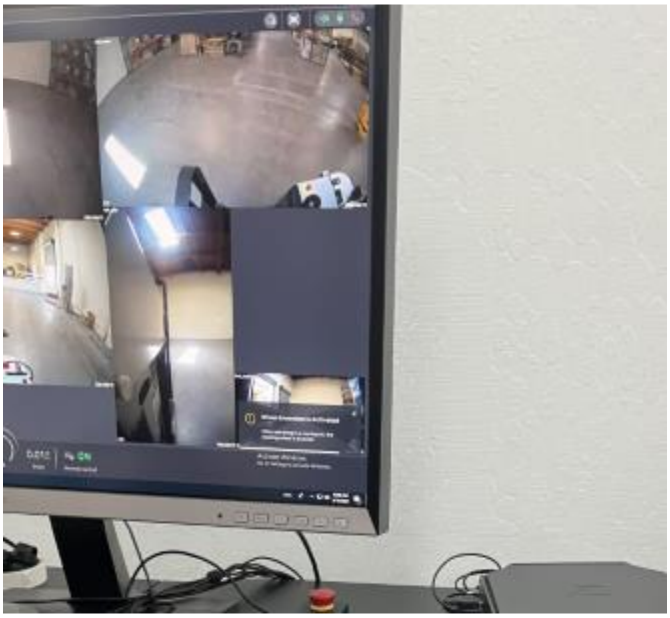


Figure 144 Popup message occluding the fork camera (bottom right of the screen)

On one occasion, wires came loose from the pallet jack housing and were hanging into the left side camera view, yet the participant didn't notice, as they were not carrying out the load or unload task and had to be asked to stop driving so the safety monitor could tuck them out of sight.

Safety elements of the interface design also presented SA barriers for participants in the flat screen condition. The fork camera was situated in the bottom right-hand corner of the screen in both monitor interface configurations. When the reverse gear was activated, a pop-up message slid into view to remind operators that the steering was inverted (the front camera was now showing the rear view) occluding the fork camera for around one second. If the participant was inserting the pallet forks to seat the load, this caused a performance delay while they waited for the pop-up to disappear (see Figure 14). In the curved screen condition, the pop-up only covered the bottom right corner of the side camera which was not detrimental to the manoeuvre and thus would not interfere with their performance time. This is a simple software change for Phantom to either move the pop-up warning to another area of the screen or further evaluating the utility of it in the future.

However, the curved screen condition also resulted in additional SA loss for participants. In the flat screen condition, they only had to move their eyes below the mast camera images to see the side cameras, whereas in the curved screen condition they needed to lean back in their chair to access the peripheral information at the sides of the screen. One participant cited the position of the side camera in the curved screen condition as causing them to operate with additional caution:

> *"I felt overly cautious at the beginning of each task and was worried about hitting objects around me. I felt frustrated by the fact that I couldn't drive faster for fear of making contact with objects."* Curved condition Participant 3

Audio loss appeared to have more impact than camera loss for participants. Usually, the first source of information that something was wrong in the warehouse came from audio feedback such as metal screeching indicating that the pallet was dragging splinters under the pallet jack rollers, or the pallet had some other type of malfunction that needed attention (a regular occurrence in warehouse logistics). Loss of audio interestingly often resulted in participants driving at higher speeds. As one participant said, "*I can't hear the damage I am doing*' when dragging a broken pallet or driving over the dock plate. With no audio feedback the operators become divorced from the remote environment despite visual information provided by the cameras, as they were more inclined to trust audio communication to give feedback on speed and pallet jack functioning.

Finally, although a key role in warehouse safety, an inexperienced safety monitor represented further SA interference as the operator had to be continually aware of where they were while they were operating. If the warehouse were unoccupied, participants could reverse confidently as pedestrian risk analysis would be unnecessary.

> *"The presence of the safety monitor was a bit stressful, specifically when they would wander around the warehouse or stand in awkward positions near narrow sections of the aisles."* Flat condition Participant 3

Forklift operators are likely to have lowest SA, and in turn have more accidents, when they are carrying out loading/unloading tasks as they are cognitively preoccupied with carrying out the necessary manouvres to complete the task (Choi et al., 2020). Even when operating remotely, drivers will still need to manage the remote work environment to ensure that there are no hazards to humans or equipment. One participant described this SA constraint when remote driving as like:

> *'having a checklist in your head… you have to do it* [carry out all safety checks] *one by one to check it's ok before you start moving'* (Participant 5 comment while operating remotely).

4.0 Discussion

4.1 Comparisons with previous research

The aim of this study was to test whether the type of monitor (Curved screen/ Flat screen) had an effect on the time it took participants to remotely operate a pallet jack to complete three tasks. Experimental data showed that the flat screen condition was associated with a performance time that was 93% of the time taken time in the curved screen condition, with the trailer unload task being the fastest task to complete and the cross dock task taking the most time. Predictably, Manual operation was quicker than remote operation, yet qualitative findings in this study indicate that remote operation confers benefits which may outweigh lower performance speeds.

Previous studies had found that curved screens at 600R resulted in the most accurate performance in visual search tasks (Kyung , 2019) and that the wider visual angles in curved screen presentations improved performance in attention-based tasks (Klatt & Smeeton, 2020). Instead, our research found that operator performance was better in the Flat screen condition. This may be, in part, due to the different layout in the flat screen condition with the side cameras directly underneath the front, left and right mast cameras. Also, the curved screen used in this study had a radius of 1800mm which is beyond the natural curve of the eye necessitating a physical change in head movement or body position to see the side cameras, whereas, in the Flat screen condition, all the cameras were directly in front of the participant. All participants stated a preference for the Flat screen condition layout in the Debrief,

> *"The flat screen being smaller in size made it easier for me to look at each camera quickly, given my preference for sitting closer to the monitor."* Flat screen condition, Participant 3,

particularly as the extra width of the monitor in the curved screen condition required more eye and head movements to look at the edges:

> *I did not like having to turn my head to see the side angle, which I use frequently."* Curved condition, Participant 4.

Surprisingly, we did not find any significant differences between the two conditions in terms of either workload or presence and task performance, in line with previous studies (Linkov & Vanžura, 2021; Schubert et al., 2001). All drivers chose either "strongly agree" or "agree" in response to questions about their subjective experience of "being there" and whether the remote environment felt 'real', which suggests the Phantom Auto interface creates a sense of telepresence in the user, but that subjective sensation is independent of the viewing display layout. Similarly, high levels of attention and visual demand were reported in both conditions which suggests that it is the task of remote operation itself which requires vigilance and visual concentration rather than specifically how the layout is presented to operators.

4.2 Limitations and modifications

We acknowledge that the confound of screen layout differences between the two conditions is a key limitation in this study, particularly as the disparity in counterbalanced presentation favoured the curved screen condition. Participants were also more familiar with the flat screen layout, so the novel presentation on the curved screen display may have affected performance times while they acclimatised to it. Participant performance steadily improved throughout the study regardless of the condition order, so further research with larger, balanced samples is advisable.

Potentially, the effect may lie in the total width of the screen rather than the curvature so further research manipulating the monitor width could provide some interesting insights into this influence. In future studies, the display layouts could be duplicated on both flat screen and curved screens by stitching all seven camera images into one 360-degree presentation and stretching the image across the full display width, whilst maintaining resolution fidelity, allowing operators to ‘turn’ the image to see behind them when required. This interface presentation could also potentially reduce the gap between "physical" and "digital" reality, as it would simulate a more authentic view of the real world on computer screens.

A key advantage of real-world studies is the degree of authenticity that they can provide, in contrast to the often narrow and unrealistic confines of many laboratory studies (Holleman, Hooge, Kemner, & Hessels, 2020). However, there is an unavoidable compromise between the strict methodological rigor that researchers can insist upon in the lab, and the unpredictable nature of the 'real world'. In this study, there were near constant interruptions to the testing process. Sometimes these stemmed from the collegiate atmosphere in the company office, as colleagues interacted with each other throughout the normal working day and occasionally did not realise that testing was underway, stopping to ask the participant (a colleague) a question. It may be better in future testing to house the study in a separate office or carry out the tests outside of normal office hours. Clients attending the office also, on occasion, coincided with participants carrying out the experimental tasks which often resulting in a crowd of interested observers and senior management behind them as they remotely operated the forklift. Many psychological researchers are concerned that the "Hawthorne effect" will positively or negatively affect performance productivity when a participant feels the gaze of an observer, yet this was such a typical occurrence during normal operation outside of testing protocols that this is unlikely to have had a confounding effect (McCambridge, Witton, & Elbourne, 2014).

The principal source of disruption and stoppages were connection issues. In the early days of the testing process, there were frequent latency spikes in at least one camera. However, this is a technology in its infancy and the concerted efforts of multiple teams working together significantly reduced the connectivity issues by the end of the testing period showing that latency, although a shared issue across all domains of teleoperation, is one that can be conquered with time.

Finally, in reality, the safety protocols which were turned off for this study would be a regulatory requirement in a working warehouse. This may mean that performance times would be slower in real life as stops may be more frequent. For example in the load/unload condition, if operators get too close to the side of the trailer it may result in a total switch off and require manual assistance. Phantom is also in development of various operating modes which will enable operations in more restrictive environments such as inside a trailer without manual assistance. This is further testament to how solutions can be found to operational challenges with advancements in technology and training.

4.3 Recommendations and future work

This study has demonstrated that whilst there is variability in how people engage with remote operation tasks, building SA whilst remotely operating a pallet jack is subject to challenges common across other remote industries, such as attention narrowing and the detrimental effect of latency loss on performance. Remote operation companies may need to enforce standard modes of operation with comprehensive pallet jack safety training, making use of the positive qualitative findings described here on the learning effects that can be achieved through observation and exposure to other operators' driving strategies. Companies should also not underestimate the downturn in remote operation execution after a period of absence from the warehouse with no opportunity for observational learning. One participant in this study, who had taken a leave of absence from the company and had therefore not operated for some months, made errors which impacted on performance time. Even trained and experienced remote operators may need to be reassessed after a break or holiday. Opportunities to observe others completing the tasks and to gain spatial knowledge of the warehouse environment may need to be considered when planning training for global workforces operating in remote locations.

Our research suggests that interface designers for ROs should take careful consideration of the scope and range of SA information shown in this study that it is possible to derive from the remote driving video display. It may be beneficial to harness the use of predictive AI software to highlight broken pallets or items on the floor which will sound an alarm to alert the RO to potential problems quickly and before they result in expensive and time-consuming damage to pallets although there will still be a need for a human to be in the warehouse for some time. Further empirical testing of the effect of the layout of cameras on-screen would narrow down whether the layout of the curved screen condition, with the side cameras positioned in the curvature at the edge of the screen, was a limiting effect on operator performance. Future studies could use eye tracking software to identify the search strategies used to access SA information necessary for each individual task, and also investigate how these might deviate between remote and manual operation.

This study has pointed to fundamental advantages in using remote operation as opposed to manual driving, such as the advances in the safety of operators in remote contexts. Advances in digital technology and increasing scale in manufacture means that there is potential to overcome the differences in SA between in situ and remote operators of forklifts, for example using variable visual feeds and visual enhancements to support driver SA. On a broader level, it is imperative that the SA needs of remote operators remain the subject of extensive, iterative research to shape operator training practices, product design and interface development in the future as automated fulfilment operations become commonplace in warehouse logistics.

References

Almeida, L., Menezes, P., & Dias, J. (2020). Interface transparency issues in teleoperation. *Applied Sciences (Switzerland)*, *10*(18), 1–32. https://doi.org/10.3390/APP10186232

Almeida, L., Patrao, B., Menezes, P., & Dias, J. (2014). Be the robot: Human embodiment in tele-operation driving tasks. *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and*

Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions, 477–482. https://doi.org/10.1109/ROMAN.2014.6926298

Chen, X., Ulmer, M. W., & Thomas, B. W. (2022). Deep Q-learning for same-day delivery with vehicles and drones. *European Journal of Operational Research*, *298*(3), 939–952. https://doi.org/10.1016/j.ejor.2021.06.021

Choi, M., Ahn, S., & Seo, J. O. (2020). VR-Based investigation of forklift operator situation awareness for preventing collision accidents. *Accident Analysis and Prevention*, *136*(January), 105404. https://doi.org/10.1016/j.aap.2019.105404

Costa, M. (2021). Frontline Forklifts. Retrieved May 17, 2022, from COLD SUPPLY CHAIN & LOGISTICS website: https://digitaledition.refrigeratedfrozenfood.com/may-2021/cold-supply-chain--logistics/

Daw, A., Hampshire, R. C., & Pender, J. (2019). *Beyond Safety Drivers: Staffing a Teleoperations System for Autonomous Vehicles*. 1–55. Retrieved from http://arxiv.org/abs/1907.12650

Endsley, M., Bolte, B., & Jones, D. (2003). *Designing for Situation Awareness: An Approach to User-Centered Design* (2nd ed.; N. Y. T. & Francis, Ed.). https://doi.org/10.1201/9780203485088

Finn, R. (2021). Gaming Monitors: How to choose the best monitors for sim racing. Retrieved from Sim Racing Cockpit website: https://simracingcockpit.com/gaming-monitors-guide/

Fryling, M. J., Johnston, C., & Hayes, L. J. (2011). Understanding Observational Learning: An Interbehavioral Approach. *The Analysis of Verbal Behavior*, *27*(1), 191–203. https://doi.org/10.1007/bf03393102

Gugerty, L. J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. *Journal of Experimental Psychology: Applied*, *3*(1), 42–66. https://doi.org/10.1037/1076-898X.3.1.42

Hart, S. G. (2006). NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER. *Proceedings of The Human Factors and Ergonomics Society 50th Annual Meeting*, 904–908.

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The ‘Real-World Approach’ and Its Problems: A Critique of the Term Ecological Validity. *Frontiers in Psychology*, *11*(April), 1–12. https://doi.org/10.3389/fpsyg.2020.00721

Iinuma, R., Kojima, Y., Onoyama, H., Fukao, T., Hattori, S., & Nonogaki, Y. (2020). Pallet handling system with an autonomous forklift for outdoor fields. *Journal of Robotics and Mechatronics*, *32*(5), 1071–1079. https://doi.org/10.20965/jrm.2020.p1071

Kennedy, Lane, Berbaum, & L. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3(3), 203-220*.

Klatt, S., & Smeeton, N. J. (2020). Immersive screens change attention width but not perception or decision-making performance in natural and basic tasks. *Applied Ergonomics*, *82*(October 2019), 102961. https://doi.org/10.1016/j.apergo.2019.102961

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Kyung, G. (2019). *Curved Versus Flat Monitors : Interactive Effects of Display Curvature Radius and Display Size on Visual Search Performance and Visual Fatigue*. https://doi.org/10.1177/0018720820922717

Law Commission of England and Wales. (2022). *Automated Vehicles : joint report*.

Lidestam, B., Eriksson, L., & Eriksson, O. (2019). Speed perception affected by field of view: Energy-based versus rhythm-based processing. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 227–241. https://doi.org/10.1016/j.trf.2019.07.016

Linkov, V., & Vanžura, M. (2021). Situation Awareness Measurement in Remotely Controlled Cars. *Frontiers in Psychology*, *12*(April). https://doi.org/10.3389/fpsyg.2021.592930

Lo, J. C., Sehic, E., Brookhuis, K. A., & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 325–338. https://doi.org/10.1016/j.trf.2016.09.006

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Michel, R. (2022). Remote lift truck operation as fleet strategy. Retrieved from Modern Materials Handling website: https://www.mmh.com/article/remote_lift_truck_operation_as_fleet_strategy

Möller, C., Kostka, A., Neujahr, H., & Engineering, H. F. (2002). *Enhancing Situational Awareness by Perspective Panoramic Information Presentation*. (May 2014), 1–7.

Niklasson, L., Riveiro, M., Johansson, F., Dahlbom, A., Falkman, G., Ziemke, T., … Kronhamn, T. (2007). A Unified Situation Analysis Model for Human and Machine Situation Awareness. *Herzog, O., Rödiger, K.-H., Ronthaler, M. & Koschke, R. (Hrsg.), Informatik 2007 – Informatik Trifft Logistik – Band 2. Bonn: Gesellschaft Für Informatik e. V.. (S. 105-109).*, *2*, 105–109.

Rosenzweig, A. (2020). Teleoperation: Safety Principles and Indirect Methods of Control. Retrieved from Ottopia website: https://ottopia.tech/blog_items/teleoperation-safety-principles-and-indirect-methods-of-control/

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Snavely, B. & Patterson, R. (2008). Do Video Display Walls Improve Situational Awareness in Control Rooms? Retrieved from Electric Energy Online website: https://electricenergyonline.com/energy/magazine/423/article/Do-Video-Display-Walls-Improve-Situational-Awareness-in-Control-Rooms-.htm

UKHaulier. (2021). UK forklift operator crisis is helping drive up sales of automated handling solutions. Retrieved from https://www.ukhaulier.co.uk/news/freight-logistics/lifting-and-handling/uk-forklift-operator-crisis-is-helping-drive-up-sales-of-automated-handling-solutions/

Witmer, B. G. &, & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Zhou, A., Wang, S., Wan, S., & Qi, L. (2020). LMM: latency-aware micro-service mashup in mobile edge computing environment. *Neural Computing and Applications*, *32*(19), 15411–15425. https://doi.org/10.1007/s00521-019-04693-w

Conclusion

The overarching aim of this thesis was to investigate the information and timescale that is required for a remote operator to gain sufficient SA to remotely control, guide or supervise an AV. In the next section, the primary aims of this research project will be examined in relation to the key findings from each Paper. Critical evaluation of the strengths and limitations of the findings are addressed and their implications for human factors research and the role of the remote operator in the future.

1.0 Key Findings, Evaluation and Implications

The research aims and objectives of this research project were achieved by the development and findings of Papers 1-6, as follows.

- Update our understanding of situation awareness in relation to remote operators of autonomous vehicles.

Paper 1 was written as a call to the field to acknowledge the significant differences between SA in normal driving contexts and SA in remote driving operations. It argued that the established understanding of automated driving required updating to include the context of remote operation that is likely to come in to play at higher levels of automation. The revision of industry standards in May 2021 added clarifications to the taxonomy and included 54 references to a “remote driver” in the new version (SAE international, 2021). Although clearly not in direct response to the article, it demonstrates the importance of publications to add pressure to the growing acceptance in the field of the role of the remote operator.

Autonomous driving applications are becoming increasingly common, for example, since the start of 2021, there has been a marked increase in the UK of trials of low-speed autonomous delivery vehicles (wilko though StreetDrone) and autonomous bus shuttles (Project CAVForth in partnership with Fusion Processing[1]), most combining autonomous technology with an RO in some capacity. The 2022 Law Commission report on the regulatory framework for automated vehicles (AVs) also recommended reforms to allow self-driving vehicles on public roads with no "user-in-charge" inside the vehicle with an RO required to oversee the vehicle and respond to requests (Law Commission of England and Wales, 2022). However, there is still a lack of clarity in the difference between remote assistance and remote driving within the higher levels of automation which can be considered a work in progress as the industry develops[2].

---

[1] https://www.fusionproc.com/

[2] https://www.sae.org/news/2021/06/sae-revises-levels-of-driving-automation

- Investigate how to effectively measure SA needs during remote operation.

Findings from Paper 2 revealed important inferences about how people build up SA from remote scenes suggesting that acquiring SA in remote scenes is a flexible process involving combining the SA Levels of 'comprehension' and 'prediction' in parallel. Previous quantitative methodologies aimed at 'measuring' each 'level' of SA and allocating it a score supposes that they are serial and linear. The qualitative methodology used in Paper 2 revealed the myriad observations, calculations and adjustments required in driving scenarios which have been overlooked in quantitative research paradigms. It was essential to develop a thorough understanding of what SA comprises in remote contexts before starting to make judgments of what "good" SA may look like and design instruments which would effectively measure SA needs during remote operation. Paper 2 generated a taxonomy of driving SA which was then used to create SA questions for the driving videos used in Papers 3, 4 and 5. Additionally, an unexpected feature of the naturalistic qualitative analysis used was the realisation that what participants *don't see* in remote driving scenes, also contributes to the building of SA, for example recognising the *absence* of hazards still constitutes careful information gathering but would not be picked up by existing measures of SA. The results of Paper 2 suggest that existing measures of SA need to be more sensitively applied to remote contexts. The taxonomy with accompanying open-source videos can also be used in future empirical work to design queries that can effectively measure RO SA.

Paper 3 used the TSA to design and validate questions designed to assess SA during remote operation. An exploratory factor analysis on the 48 items from the SA measure found four underlying constructs which we termed 'spatial and environmental awareness', 'anticipatory hazards', 'dynamic driving actions' and 'other road users'. Although the findings of all the studies in this thesis generally support Endsley's (1988) concept of SA as incorporating what is known about the environment, what is happening in it and what might change to build awareness, results from the factor analysis suggests that focusing too narrowly on the divisions between the three levels of SA may lack construct validity in the context of driving SA. This assumption was echoed in Paper 4 which applied the SA questions in an empirical study, finding low SA performance overall and no significant difference in performance between the SA levels of perception, comprehension and prediction. One implication of these findings is that measurement tools designed to hold SA levels separate may be inappropriate in the context of SA needs during remote operation. Paper 5 also provided support for the factor structure outlined in Paper 3, as different categories of edge case sceanrios represented by the 60 videos used in the study were consistent with the underlying dimensions of the SA measure used in Papers 3 and 4. This advocates the inclusion of different types of edge cases in any design of a new measure of remote SA as they can reasonably be assumed to fit within the underlying constructs that Study 3 indicated were significant. However, the sample size in Paper 3 was below that recommended to conduct EFA so conclusions about the nature of remote SA from this measure must be taken with caution and further replication with larger samples would be advised. However, the findings from Papers 3, 4 and 5 represent a strong contribution both towards developing a more sensitive instrument to effectively measure driving SA in remote environments and further understanding of the underlying principles of remote driving SA.

- Contribute to the understanding of the optimal format information should be delivered to an RO to build and maintain SA quickly and safely in different vehicle remote operation scenarios.

Paper 4 investigated whether the provision of rear-view and/or auditory information could improve SA for remote driving scenes, presenting the videos from the perspective of the visual feed of a remote operator to understand the optimal format information should be delivered to an RO in order to build and maintain SA quickly and safely. The initial hypothesis was that SA performance would be highest in the condition in which both the rear-view and audio feeds were presented, as this would provide the most information to assist in building awareness. The finding that SA performance was worse in the presence (vs. absence) of the rear-view feed with no effect of audio was surprising as research (Ho & Spence, 2005; Störmer, 2019) had indicated that audio feeds would be helpful for directing attention. Instead, it would appear in this study, that participants found it distracting. Furthermore, in this task, participants may have quickly worked out that the information in the rear-view feed was unnecessary as none of the SA questions asked about information contained within the rear-view feed. In which case, it may have represented 'better' SA to ignore the rear-view feed. These findings suggest that designers of interfaces to support an RO developing and maintaining SA may need to consider carefully how they present rear-view and audio feeds, possibly depending on task requirement.

Paper 5 investigated the optimal format information should be delivered to an RO by testing the effect of HMD-360 presentation or Screen-based 360° presentation when carrying out a remote supervision task. The results suggested that decisions were made in the HMD-360 condition in less time than that required in the SB condition and corresponded with more accurate decisions. Commonly, concerns about cyber sickness symptoms resulting from HMD-360 presentation have discouraged companies from engaging this form of technology in remote operations, but in this study, although participants reported that they found the headset uncomfortable there were no cyber sickness symptoms reported in either condition. This is probably because the participant themselves were not in motion and thus did not have conflicting vestibular information which frequently results in cyber sickness. This suggests that HMD-360 may be more suitable for operators carrying out a supervisory role in fleet operations rather than active driving control of an AV. The findings from Paper 5 advocate that companies offering remote supervision of AVs should explore the potential of HMD-360 presentation to enhance operators' remote SA but limit its use to shorter periods of operation to reduce negative workload issues. Exploratory analyses in Paper 5 also revealed that decisions in response to videos which required highway code comprehension took less time than was taken than other videos, proposing that rehearsing a 'Highway Code' equivalent during RO training might help to ensure that they are making timely and accurate responses to AV requests for assistance. Eye tracking analyses demonstrated differences in the duration of fixations in AOIs defined in the video image depending on whether it was presented in SB or HMD-360 format, together with a left hand bias in the SB condition and a right hand bias in the HMD-360 condition.

Paper 6 was a real-world study with Phantom Auto investigating whether the type of monitor (Curved screen/ Flat screen) influenced the time it took participants to remotely operate a pallet jack to complete three tasks. Results showed that the flat screen condition was associated with quicker performance times than the time taken time in the curved screen condition, with the trailer unload task being the fastest task to complete and the cross-dock task taking the most time. The confound of screen layout differences between the two conditions was a key limitation in this study, particularly as the disparity in counterbalanced presentation favoured the curved screen condition and participants were also more familiar with the flat screen layout. It is not possible to draw meaningful conclusions as to the suitability of either screen type for remote operation of forklifts from this study, however in-depth qualitative comparisons between remote and manual operations in this study endorsed the potential of digital technology to overcome the differences in SA between in situ and remote operators of forklifts. Remote operation in warehouse logistics may confer many benefits which may outweigh lower performance speeds with the proviso that further research into designing interfaces that have been empirically proven to support remote SA is needed.

- Investigate factors that will increase an RO's sense of presence in the remote scene.

Papers 4, 5 and 6 all included measures to assess participants' sense of presence in the remote scenes. Paper 4 found the highest sense of presence in the conditions which had rear-view information present but overall, the scores were all low in terms how much of a sense of presence the driving videos produced. Paper 6 found no difference between perceived 'realness' or sense of 'being there 'in the flat screen condition and the curved condition although scores were high in both. These results contradict consistent findings that participants rate their sense of presence higher in conditions which provided additionally sensory information such as sound and extra or enhanced visual feeds (Cummings & Bailenson, 2016; Georg et al., 2020; Larsson, Västfjäll, Olsson, & Kleiner, 2007; Schubert, Friedmann, & Regenbrecht, 2001). Potentially, this could be due to the small number of data points in data collection (only one for each participant) which may mean that the study was not sufficiently powered in this question to detect an effect. However, the low presence ratings in Paper 4 may be best explained by the passive nature of the task, as participants may have felt detached from the images they were watching. Conversely, the high presence scores in both flat and curved screen conditions in Paper 6 could be due to the fact that participants were directly operating the pallet jack in the task/s. Conceivably, it is being active in the remote scene, rather than an observer, that has the most impact on the sense of presence rather than how the layout is presented to the remote operator. This is supported by the significant presence results in Paper 5, with the HMD-360 condition conferring the highest sense of presence, as participants had to view the videos and make a decision (resulting in a button press), so they felt involved in the remote scene. Findings across these studies imply that the mode of presentation in remote contexts influences a sense of presence but, critically, being actively involved with events within the remote scene is also associated with feeling ‘there’.

2.0 Critical evaluation of the experimental task

In many of the experimental tasks conducted in this PhD thesis, participants were not actively driving and so were not subject to attention being diverted by other driving tasks such as changing gears, depressing pedals and other peripheral distractions (e.g., changing music). However, full teleoperation is unlikely to be the most frequent task that an RO will carry out, with remote supervision likely to be more common so this is not a diminishing factor in any of the thesis studies. Papers 2, 3 and 4 asked participants to watch videos of driving relay and answer SA questions on the content of the videos. This simplified task does not capture an authentic RO experience of dropping-in unexpectedly to a remote scene and having to acquire SA as, in reality, an AV is required to assume a fallback state of the minimal risk condition (MRC), a safe stop, if there is a problem that requires an RO. However, these studies enabled the understanding of how remote SA builds up over time, so viewing dynamic driving from a remote perspective was appropriate during the genesis of the qualitative information necessary to construct the TSA and testing of the resulting SA measure. Findings from these tasks facilitated valuable insights about the range and scope of information that can be derived from remote driving videos and the variability between participants in how they construct SA of remote scenes.

Paper 5 addressed this lack of realism in previous experimental tasks, by reflecting the MRC in its experimental task; deciding the next direction the AV should take, based on the information presented in the remote scene from the persepctive of a stationary AV. Paper 6 was a real-world study where operators were moving pallets remotely in a real warehouse so reflected authentic remote operation. The findings from these papers confirm findings from the experimental studies which had less mundane realism – namely that visual feedback and format of information presentation to the operator significantly impact on the ability to develop and maintain SA in remote contexts – providing support for the integrity of the overall research process across the PhD thesis. Interestingly, findings were conflicting across studies in the thesis in terms of the usefulness of audio information; although the audio feed did not appear to be helpful when watching driving videos (in Paper 4), it was a critical requirement during real teleoperation (in Paper 6). This demonstrates that validating empirical research findings through real-world applications will be an important consideration in transportation research in the coming years.

3.0 SA needs during remote operation will differ to SA requirements during in-situ driving

While the focus of this thesis was situation awareness of remote vehicle operators, some of what was discovered during its course is likely to apply to in-situ driving too. For example, there are common SA 'demons' discussed in Paper 1, such as attention tunnelling, cognitive overload and out of the loop (OOTL) syndrome, combined with embodiment and presence issues and workload, which will all factor in both remote and in-situ driving. Furthermore, the individual differences between participants in how they collect SA, observed in Paper 2, is also likely to apply to driving in-situ. We might expect remote operators to be more highly trained and drawn from experienced labour pools, which will likely influence their SA collection techniques, for example experienced drivers have been found to use their exterior mirrors more than new drivers (Crundall, 2002). However, Paper 6 demonstrated that operators will still have individualised approaches to carrying out both in-situ and remote operational tasks. Similarly, the interplay between different levels of SA described in Paper 2, where we found evidence that a "higher" level can feed downward to a "lower" level (meaning that SA Comprehension or SA Prediction may influence perception and vice versa), may not be confined to remote contexts but, instead, is more typical of 'driving SA'. Visual information processing during driving is possibly more complex than that required for pilot SA which was the focus of Endsley's original model, but remote driving and in-situ driving are still similar in terms of awareness needs (Endsley, 1988).

There are subtle differences however between the information received when occupying a vehicle in-situ and remotely operating through video feed, particularly in the context of the sub-theme 'Rationale/analysis of other driver's action (theory of mind)' outlined in Paper 2. When interactions with other road users are viewed through a visual feed an RO could plausibly miss subtle micro gestures. For example, when filming the driving videos used in Paper 4, I ensured that I kept rigidly to speed restrictions to produce ‘perfect’ examples of driving. Despite this (or probably because of it!) on more than one occasion, other drivers swore at me, using rude hand gestures from inside their cars, which I could easily see through my window sitting in-situ, yet their behaviour was not visible in the driving videos. Hence, some visual information transmitted by monitor feed to an RO is restricted compared to sitting inside the vehicle developing good SA more challenging. If remote operators have limited interactions with other road users or are unaware of the effect they may be having on them, this could have serious consequences in public acceptance of highly automated cars.

Use case is critical when judging whether SA needs will be worsened or improved for ROs. The range of the functional scope of a remote operator, outlined in Paper 1, covers ‘Remote assistance’, ‘Remote supervision’ and ‘Remote control’, and each role may potentially be different across industries with uniquely different respective SA needs. For example, ‘Remote control’ involves remote driving; a continuous, dynamic task meaning that the RO can maintain and update their SA throughout the task, like in-situ driving. Yet degraded visual information and possible latency is unique to remote driving and subsequent turning, accelerating, and braking could all be delayed if the image the RO is seeing is out of date in terms of useful interaction (T Systems, 2020). In Paper 6, during a sizeable latency spike of 3-4 seconds in the front mast screen during remote operation in the ‘Curved’ condition, the participant did not see that the other cameras were showing him driving towards the warehouse racks. Without a safety system (lidars or other technologies) to prevent impacts, the latency of the primary video feed may have resulted in a crash. A comparable in-situ example could be a driver sneezing momentarily, then opening their eyes and continuing driving as the world around them has hardly changed in those split seconds. Supplementary camera feeds can provide ROs with enough information to maintain SA and continue driving, if they use those visual feeds instead, but only if operators are confident that complete video feeds will be quickly restored.

The use case of ‘Remote supervision’ involves more variance in the SA challenges between an RO and in-situ driver. If an AV requests assistance, an RO has had no time to build up SA prior to a ‘drop in’ as they have not been monitoring the situation beforehand. The task utilised in Paper 5 was designed to simulate a supervisory decision-based task with no prior warning. Results showed that the format of presentation of the visual information was important, with HMD-360 presentation reducing the time taken to make a decision, but it still took participants between 2-12 seconds, dependent on the type of edge case represented by the video and the presentation format (SB or HMD-360). Evidence from Paper 6 implies that reduction in visual feeds is exacerbated if audio cues are also lost during remote operation of warehouse logistics – although this may be less noticeable in remote supervision tasks where audio outside the vehicle may not be available. Thus, building (and maintaining) SA in the use case of 'Remote supervision' will take longer than for an in-situ driver located inside the vehicle and at the local environment. Physical absence makes the task of developing SA from a remote location more difficult; accordingly, SA needs for an RO are more demanding compared to those of an in-situ driver. As argued in Paper 3, standard measurements to assess SA in driving may not be appropriate for remote operation as the scenario of a remote 'drop in' to an AV strongly suggests that SA needs during remote supervision will significantly differ to SA requirements during in-situ driving.

A further use case where SA needs will be different to in-situ driving is fleet operation management. Transport companies, such as Oxbotica[3], have predicted centralised control hubs which AVs can contact when they encounter an edge case. Managing a remote fleet will have similarities with situated fleet management, such as carrying out vehicle maintenance and location tracking of vehicles but has the potential to be even more efficient by using telematics and advanced software systems to stay connected[4]. However, understanding the unique SA demands on ROs workload in respect of allocation to jobs, prioritising AV requests, accidents involving other road users and managing shift changes will be critical in ensuring the safety of remote fleet management and will be unique to remote operation (Cummings, Li, Seth, & Seong, 2020).

Furthermore, different industries will have diverse RO SA needs. For example, for companies which offer ride hailing mobility solutions, such as Imperium Drive, the RO role will plausibly involve remote supervision as well as remote driving, but within warehouse logistics, for example, forklift operation offered by Phantom Auto, remote driving will be non-stop due to current teleoperation requirements. As discussed above, each use case of an RO carries more or less of an SA burden. Different forms of autonomy will also require more or less intervention than others and will therefore impose different levels of workload on ROs. For example, remote operation of 'last mile' deliveries in a local environment compared to long distance trucking creates different demands on RO SA (Mutzenich et al., 2021).

It must be acknowledged however, that the extra information which can be accessed via advanced software and telematics means there is potential for SA to be improved for remote driving

---

[3] https://www.oxbotica.com/our-technology/
[4] https://envuetelematics.com/

compared to in-situ. For example, in Paper 6, the camera interface presentation in the 'Flat' screen condition positioned the side cameras directly under the mast cameras making it easier to change viewing perspective when remotely operating than when manually driving. The range of cameras and additional sensory information provided by an AV may give an RO enhanced awareness of some aspects of the environment or draw attention to information that could be neglected in person. In the HMD-360 condition in Paper 5, participants found it difficult to turn all the way around in the chair to access details in the 360° video image directly behind them, so using technological feeds to inspect a remote scene may be easier than visual perception in-situ.

In sum, when assessing the unique SA needs for ROs, use case and industry matters. Cross Parliamentary groups in the UK are already recognising the need for a "flexible regulatory framework that allows use cases to advance and innovate"[5]. This is sensible and realistic to facilitate adoption of highly automated transport solutions, provided a framework is set for understanding SA in driving separately to SA in other industries. It is essential for future research to develop a broader definition of remote SA that can be applied in a nuanced way to each distinct use case of an RO.

4.0 Directions for future work

There are several projects[6] currently underway in the UK designed to understand a range of aspects of remote operation, including human factors considerations, with the aim of enabling remote operation to become a feasible transport opportunity. In normal driving contexts, drivers maintain SA by distributing their gaze across the road to identify hazards, signs and carry out tactical manoeuvres (Merat et al., 2019). Future research could make further use of the extensive eye tracking data collected in Paper 5 for more advanced analysis, such as using scan path analysis to plot how gaze points evolve during the viewing of a remote scene (e.g., in what order people move their eyes around the scene to collect SA information). It is possible that the pattern of gaze distribution would differ for ROs of AVs in different use cases, for example those playing a supervisory role may display gaze patterns more akin to watching TV than ROs performing an in-situ dynamic driving task, due to the nature of how the visual information is transmitted to them on a computer monitor. Eye tracking studies labelling the dynamic eye movement of experienced operators driving remotely could help to understand how SA is constructed in dynamic interactive tasks. Further research directions based on the findings of this PhD project point towards testing the

---

[5] https://appgconnectedandautomatedmobility.car.blog/

[6] https://www.gov.uk/government/collections/connected-and-autonomous-vehicles-projects-case-studies

effect on remote SA of the inclusion of spatial audio in the design of remote interfaces remote contexts to enhance attention and direct visual perception.

Key questions for remote operation that were outside the scope of this PhD project, but are still critical within the automated transport field are,

- **What are effective SA strategies in remote driving?**

  This PhD thesis has demonstrated that there are important differences in building and maintaining SA in remote contexts compared to in-situ drivers, so this is likely to also be the case when considering the strategies that drivers adopt when taking active control of the dynamic driving process. Future research should compare steering manoeuvres, accelerating and braking forces and driver performance on a simple driving route carried out under in-situ and remote operation conditions to determine effective remote operation strategies.

- **What is the optimal interface layout or mode of presentation for remote operators?**

  The type of information that will be useful for ROs to construct SA of a remote scene varies considerably together with how that information is presented. Future research should investigate how interface presentation could be leveraged to reduce the gap between "physical" and "digital" reality, for example by stitching images together taken from the cameras around the AV to simulate a more authentic view of the real world on computer screens.

- **How should we manage handovers between remote drivers or vehicles and what are the SA implications involved?**

  There are autonomous transport companies already operating, such as Einride , who envisage remote operators monitoring and controlling multiple autonomous vehicles from a single remote drive station. The cognitive cost of switching between remote locations at potentially a flick of a switch will have significant implications for ROs having to repeatedly build and re-build SA and it has crucial implications for safety (Cummings, Li, Seth, & Seong, 2020). Further research is recommended to consider how to prioritise requests for assistance from multiple AVs, which may arrive before an operator has finished gaining the necessary SA to act in the first event.

5.0 Contribution to the Field

This thesis offers an original theoretical contribution of a new consideration of how to measure SA in remote driving. Although acknowledging Endsley's Levels 1/2/3 as the core basis of how SA is constructed, I have argued throughout this work that a novel approach is needed in order to encompass the supervisory role that an RO may play. An RO is unlikely to have been monitoring the scene and may need to build understanding of the goals of the AV from scratch, which suggests that SA needs during remote operation will differ to SA requirements during in-vehicle driving. Hence, standard measurements to assess SA may not be appropriate. Taken together with the lack of a unified approach to SAGAT methodology (beyond the use of a freeze and probe approach), discussed in detail in this work, these weaknesses point to a need for new methods to measure driving SA and to adapt them across RO use cases as required. Increasingly, transportation research is concentrating on questionnaire development and behavioural measures to understand driving, and this thesis has made a strong step towards developing a more sensitive instrument applied to measuring driving SA in remote environments. However, there is much work to be done in this arena. We should identify the skills we wish ROs to develop and design training videos and accompanying questions to extract those aspects, rather than relying on all-purpose approaches across many domains, such as SAGAT.

This thesis also makes an original methodological contribution in the approach taken to transform the screen based eye tracking into corresponding X,Y coordinates in the video stimuli. Screen based eye trackers can tell you where on the screen the participant is looking but not ‘what’ they were looking at in a video playing on the monitor. Many studies have performed screen-based eye tracking on images, videos interfaces etc. to provide deeper insights into visual attention but the innovative process of transformation for the X,Y coordinates in the SB condition in Paper 5 can now be replicated in future studies which also wish to use video-based stimuli and screen based eye tracking to measure SA. A further contribution of the thesis is to provide a complete, open-source set of videos of remote driving scenes and AV edge case scenarios[7] which can be used in future research projects investigating remote operation.

This thesis makes many practical contributions to the connected and automated vehicle industry. It calls attention to evidence that the task of gaining and maintaining SA for ROs will depend heavily on the RO use case, either direct remote control or supervisory tasks and the segment of the transportation industry that they are working in. From personal experience, when designing and producing the video stimuli for the studies in this thesis, I would advise caution in making assumptions people will always ‘see’ the same information in a scene, as people often identified details that even I had missed! The individual variability when developing SA must be taken into account when designing training protocols for ROs in future workplaces. Furthermore, extra information is not always desirable and may occasionally impair performance; this should warn designers of interfaces

---

[7] https://github.com/Anthrometric/remote-operator.git

for ROs that each piece of information that is made available to ROs should be empirically proven to assist in the construction of SA in remote contexts.

Although I encourage companies offering highly automated mobility solutions to consider new technologies such as HMD-360 to increase presence and speed up decision making, some findings within this thesis suggest that it would be most appropriate to limit HMD-360 to shorter periods of operation to reduce negative workload issues. HMD is probably more suitable for operators carrying out a supervisory role in fleet operations, where the nature of the engagement is likely to be briefer, than for operators engaging in extended periods of teleoperation. It will be interesting to observe, as the industry moves closer towards deployment, whether legal and regulatory bodies will legislate certain modes of presentation when remotely operating based on evidence as to whether they aid or impede operator SA.

A safety consideration raised in this work is that, in the future, on-road vehicles in the UK may be being remotely driven from abroad requiring switches between left- and right-hand driving. The Law Commission has raised the question in a public consultation of whether this has the potential to create serious practical and legal difficulties in enforcement[8]. However, findings in this thesis point towards significant endangerments concerning where remote operators will look in the remote scene for information to build SA, dependent on how the information is presented to them. Visual search strategies in remote visual feeds may be susceptible to left-hand bias if video feeds are presented via monitor display, compared to right-hand bias if viewed in an HMD, which may influence both time to come to a decision in a supervisory task and the accuracy of that decision. These findings could have repercussions within the on-road transportation industry when remotely operating on-road vehicles in left or right-hand driving countries. Until further research is conducted into the cognitive burden of switching between sides of the road when remotely operating and the optimal mode of presentation to operators to safeguard against errors of judgment and produce optimal SA, I recommend that driving from abroad should be prohibited when AVs are finally deployed.

Finally, some commentators in the transport field (often in blogs and other thought leadership pieces) believe that the AVs should not be on public roads without a safety driver until the vehicle can handle every possible emergency, including novel situations, with the same capability of a proficient human driver. Instead, I recommend widespread adoption of the RO role to support AVs. As a human will be on call and ‘in the loop’ to assist the AV if it is in difficulty, this move could bring automated transport solutions and all the resulting benefits a step closer sooner.

---

[8] https://www.lawcom.gov.uk/law-commission-starts-debate-on-how-to-regulate-remote-driving/

References

Adam, E. C. (1993). Fighter cockpits of the future. *[1993 Proceedings] AIAA/IEEE Digital Avionics Systems Conference*, 318–323. https://doi.org/10.1109/DASC.1993.283529

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorilla in our Midst: An online behavioral experiment builder. *BioRxiv*, (April 2019), 438242. https://doi.org/10.1101/438242

Braun, V., & Clarke, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*, *26*(2).

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with some simple guidelines. *Journal of Cognition*, *2*(1), 1–38.

Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Cummings, M. L., Li, S., Seth, D., & Seong, M. (2020). *Concepts of Operations for Autonomous Vehicle Dispatch Operations*. 32. Retrieved from https://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u36/CONOPS dispatch_Final_compressed.pdf

Davies, A. (2019). The War to Remotely Control Self-Driving Cars Heats Up. Retrieved October 13, 2020, from Wired website: https://www.wired.com/story/designated-driver-teleoperations-self-driving-cars/

Endsley, M. R. (1988a). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting*, *32*(2), 97–101. https://doi.org/10.1177/154193128803200221

Endsley, M. R. (1988b). Situation awareness global assessment technique (SAGAT). *Aerospace and Electronics Conference*, 789–795. https://doi.org/10.1109/NAECON.1988.195097

Georg, J. M., Putz, E., & Diermeyer, F. (2020). Longtime Effects of Videoquality, Videocanvases and Displays on Situation Awareness during Teleoperation of Automated Vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *2020-Octob*, 248–255. https://doi.org/10.1109/SMC42975.2020.9283364

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied*, *11*(3), 157–174. https://doi.org/10.1037/1076-898X.11.3.157

Jewitt, C. (2012). *National Centre for Research Methods Working Paper: An introduction to using video for research*.

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE Life Sciences Education*, *18*(1), 1–17. https://doi.org/10.1187/cbe.18-04-0064

Larsson, P., Västfjäll, D., Olsson, P., & Kleiner, M. (2007). *When What You Hear is What You See : Presence and Auditory-Visual Integration in Virtual Environments Applied Acoustics , Chalmers University of Technology , Göteborg , Sweden , 2 Department of*. 11–18.

Law Commission of England and Wales. (2022). *Automated Vehicles : joint report*.

Ledesma, R. D., Ferrando, P. J., Trógolo, M. A., Poó, F. M., Tosi, J. D., & Castro, C. (2021). Exploratory factor analysis in transportation research: Current practices and recommendations. *Transportation Research Part F: Traffic Psychology and Behaviour*, *78*, 340–352. https://doi.org/10.1016/j.trf.2021.02.021

Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, *23*(6), 736–757. https://doi.org/10.1080/13506285.2015.1079583

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, *6*(JAN), 1–7. https://doi.org/10.3389/fpsyg.2015.00002

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., … Keinath, A. (2019). The “Out-of-the-Loop” concept in automated driving: proposed definition, measures and implications. *Cognition, Technology and Work*, *21*(1), 87–98. https://doi.org/10.1007/s10111-018-0525-8

SAE international. (2021). *SAE J3016; Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Störmer, V. S. (2019). Orienting spatial attention to sounds enhances visual processing. *Current Opinion in Psychology*, *29*, 193–198. https://doi.org/10.1016/j.copsyc.2019.03.010

Watkins, M. W. (2018). Exploratory Factor Analysis : A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807