Investigating the Effects of Augmented Reality Cues During Non-Driving Related Tasks on the Situational Awareness of Drivers

Thomas Alexander Goodge

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Computing Science & School of Psychology and Neuroscience
College of Science and Engineering
University of Glasgow

February 2025

Abstract

As automated vehicles become responsible for more of the driving task, the way in which drivers need to process the road will change. This will free up capacity to engage with non-driving related tasks, activities which are unrelated to operation of the vehicle or attending to the road. However, the current state of automation still requires a driver to be ready to resume control of the vehicle at all times. This thesis evaluates the effects of using augmented reality for presenting these non-driving tasks to drivers of automated vehicles. In particular, the ability of drivers to react to a hazard and predict what happens next in a road scene, a key component of situational awareness, were measured while they were performing a non-driving related task presented in augmented reality. Six experiments, using a mix of validated empirical tests of situational awareness, expert focus groups and eye tracking measures, were designed and conducted to assess the impact of engaging with a distracting non-driving task while attempting to maintain attention on the road. Results showed that, contrary to prior recommendations, a heads-up display presentation of a non-driving task at eye level does not convey the same benefits found when displaying non-driving related information. Evidence of intentional blindness was found when evaluating the use of attentional cues within a dynamic augmented reality display. This demonstrated that using eyes-on-road as a measure of attention is not wholly appropriate when investigating how an augmented reality interface overlaid onto the real-world impacts driver attention. Further exploration into how to design efficacious attentional cues highlights how the inclusion of salient attention capturing elements in a positional cue can enhance driver awareness of the road when they are distracted by an NDRT. Additionally, it was shown that drivers were able to utilise social cues from a virtual agent highlighting the position of a hazard, indicating the potential application of this modality for enhancing situational awareness through in-vehicle virtual assistants. This thesis contributes to the field by providing evidence of the impact of presenting an NDRT via AR on driving performance measures and methods in which this can be overcome with attentional cues. Overall, the findings in this thesis have significant implications for the applied transport psychology and automotive user interfaces domains.

Contents

List of Tables

List of Figures

Abbreviations

ANOVA	Analysis of Variance
AOI	Area of Interest
AR	Augmented Reality
ARETT	Augmented Reality Eye Tracking Toolkit
AV	Automated Vehicle
CI	Confidence Interval
DBQ	Driver Behaviour Questionnaire
Est	(Model) Estimate
GLME	Generalised Linear Mixed Effects (Model)
HCI	Human-Computer Interaction
HMI	Human-Machine Interfaces
HRI	Human-Robot interaction
HP	Hazard Perception
HUD	Heads-Up Display
HDD	Heads-Down Display
Hz	Hertz
IVIA	In Vehicle Intelligent Agent
LHD	Left Hand Drive
LME	Linear Mixed Effects (Model)
ML	Maximum Likelihood
MRTK	Mixed Reality Toolkit
MS	Milliseconds
NASA TLX	NASA Task Load Index
NDRT	Non Driving Related Task
REML	Restricted Maximum Likelihood
RHD	Right Hand Drive
RoSAS	Robotic Social Attribute Scale
SA	Situational Awareness
SD	Standard Deviation
SAGAT	Situational Awareness Global Assessment Tool
SSQ	Simulator Sickness Questionnaire
TOR	Takeover Request
UI	User Interface
VR	Virtual Reality
WHN	What Happens Next (Hazard Prediction test)

Ethics

Ethical approval for the experiments presented in this thesis was granted by the University of Glasgow College of Science and Engineering Research Ethics committee and resulting data was processed in line with legal requirements. The associated ethics application reference numbers for these experiments are:

Chapter 3:	Experiment 1	#300210133	Experiment 2	#300210313
Chapter 4:	Validation	#300220138		
Chapter 5:	Experiment 3	#300220199	Experiment 4	#300220198
Chapter 6:	Experiment 5	#300230057	Experiment 6	#300230120

Acknowledgements

First, I owe my deepest thanks to my supervisors, Stephen Brewster and Frank Pollick, who have patiently suffered through early morning meetings and my ramblings for the last 4 years. Your expertise, encouragement, and trust in my abilities has allowed me countless opportunities to grow as a researcher and produce this thesis. Thank you to the Social AI CDT for giving me this PhD opportunity after what felt like a terrible interview! My thanks go to Alessandro and Monika for your leadership over and support over the years, and to Jared, Monika, and Stuart for always being there to answer my endless questions and helping me deal with the big university admin machine. Moving to Glasgow to start a PhD during the height of COVID in 2020 was a real challenge, and I would not have made it through without the support of my peers in Cohort 2. Special thanks to Morgan for our lockdown walks and shared struggles as "non-technical" students, and to Gordon for your endless patience while teaching us all Python.

I've been incredibly lucky to belong to two research groups during my time at Glasgow. My thanks to the Multimodal Interaction Group and the Perception, Action, and Cognition Group for providing welcoming, collaborative and friendly research environments. Special mentions go to Paddy for being an unofficial PhD older sister, always with sound advice and papers that I needed to read, Kat for being such an inspiring and motivating researcher, Gang for your boundless enthusiasm for my research and unstoppable recruitment drive, Melvin, we went from clueless first years to slightly less clueless third years, your brilliance has always inspired me, Kieran for guiding me through CHI and showing me an incredible work ethic, and Ammar for your infectious enthusiasm, passion for research and selfless support of others that was incredibly inspiring to share in, you are a real Silver Star. Thank you as well to the Interactive Media Design Lab at NAIST for tolerating me while I was writing this thesis, and to Sawabe-先生 for inviting me to Japan: ありがとうございます! I'd also like to thank my original research group from Nottingham Trent University, particularly David, Petya, Andrew, and Editha, for allowing me to work on such interesting and innovative projects that introduced me to transport psychology. In particular, thanks to Vikki: you've been such a role model for me from day one, and I owe so much of my development as a researcher to your mentorship.

I would also like to give special thanks to: my teaching partner Serena, for your inhuman ability to tackle problems and sharing your endless optimism with me; my supervisor twin Jackie, for your invaluable friendship and counsel over the years; my brilliant flatmate Tobi, for your resilience, humour, and unending willingness to help out; and my anchor Emily, for guiding me through some of the toughest times of this PhD and always being there to listen and quell my needless worries. Your support has been vital and I would not have made it this far without you. Finally, to my family for always being there from start to finish, regardless of whether it was sharing my success or consoling my failures. Thank you to my Mum and Dad for not only reading through this incredibly long and onerous document, but also for doing so with encouragement and without complaint, just as you have supported me always.

Declaration

The research presented in this thesis is entirely the author's own work. This thesis only presents the parts of the paper listed below that are directly attributable to the author:

Experiments 3 and 4 in Chapter 5 have been published and presented at CHI 2024: Thomas Alexander Goodge, Frank Pollick, and Stephen Anthony Brewster. 2024. *Can You Hazard a Guess?: Evaluating the Effect of Augmented Reality Cues on Driver Hazard Prediction.* In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 254, 1–28. https://doi.org/10.1145/3613904.3642300

All data files and data analysis scripts can be found at the github repository:
https://github.com/thomasgoodge/PhDRScripts

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution. I hereby give my permission for this thesis to be shown to other University of Glasgow students and to be distributed in an electronic format.

Thomas Alexander Goodge

"I don't think this was rocket science, but I think it's a valid paper which can contribute to future road traffic safety."

Reviewer 6

Review comment on the first academic paper I ever submitted, which is potentially still relevant...

Chapter 1

General Introduction

The sophistication of Automated Vehicles - (AVs) has increased dramatically in recent years (Faisal et al., 2019). Governments around the world are starting to allow the introduction of AVs onto the road, such as in the USA (U.S. Department of Transportation, 2021), the EU (The European Parliament, 2022) and the UK (UK Government, 2024). Despite it being a common term, 'autonomous' driving encompasses a wide array of different and specific technologies. The Society of Automotive Engineers (2018) set out 5 definitive levels of automated driving, ranging from full manual control of the vehicle up to full automation, where there is no need for human intervention in the driving task. However, despite the promise of fully replacing drivers with computers, the progression of automated vehicle technology to this level has not progressed as fast as anticipated (Hancock, 2019). The partially automated systems that are currently available still require a human operator to maintain supervision of the driving task as a failsafe, due to operational constraints of the automated system. The driver-supervisor is required to stay attentive to the road at all times and must be ready to take over control. Yet, the promise of taking away responsibility for the driving task is reported to be the main motivation for consumers to own an automated vehicle (Le Vine et al., 2015; Panagiotopoulos and Dimitrakopoulos, 2018). Furthermore, supervision itself is not a task that humans are cognitively predisposed to (Parasuraman and Riley, 1997; Kyriakidis et al., 2019). Bainbridge (1983) demonstrated the deterioration of attention and cognitive performance when a task changes from manual to a supervisory in an industrial setting. Studies investigating sustained attention to automated processes indicate a marked drop in cognitive performance as the time spent supervising increases (Head and Helton, 2014; Esterman and Rothlein, 2019). As AVs start to replace human-driven vehicles, drivers will be required to perform supervision tasks they are neither predisposed to nor trained for (Kovács et al., 2021).

In addition to this, drivers do not *want* to spend their time supervising an AV. Rather, they report wanting to be able to relinquish control of the driving task to engage with Non-Driving Related Tasks - (NDRTs), such as socialising with passengers, reading, playing games, browsing the internet or being productive (Panagiotopoulos and Dimitrakopoulos, 2018; Merfeld et al., 2019). This has been illegal up until now in manually driven vehicles, due to the overwhelming evidence of the impact performing a distracting secondary task has on driving performance (Young et al., 2007; Ige et al., 2016; Lipovac et al., 2017). A similar pattern is evident in AVs, where resuming control of the vehicle after supervision is impaired by secondary tasks (Merat et al., 2012; Pipkorn et al., 2022). A meta-analysis of studies from Zhang et al. (2019) identified performing an NDRT on a handheld device as particularly detrimental on the time it takes to resume control of an AV.

Consequently, a joint report from the UK & Scottish Law Commissions (UK and Scottish Law Commissions, 2020) recommend that NDRTs should only be permissible in AVs if they *"do not prevent the driver from responding to demands from the automated driving system"*. The resolution states that the user should be *"ready and able to take control"* and *"maintains the capabilities necessary to fulfil their respective duties"* (UK and Scottish Law Commissions, 2020). In line with this guidance, governments are starting to introduce legislation stating that drivers in partially automated vehicles are responsible for maintaining attention to the road and must be ready to resume control at a moment's notice (UK Parliament, 2024; UK Government, 2022b; The European Parliament, 2022).

1.1 Motivation

This presents a conflict between the goals and motivations of drivers to use an AV and the legal, infrastructural, and technological limits on automated vehicles that are currently available. Given that more sophisticated automated driving technology which removes the need for a human is not projected to become commercially available for many years (Kosuru and Venkitaraman, 2023), it is important to consider how to allow drivers to engage with an NDRT while still being responsible for supervising the driving task.

In an attempt to mitigate this, much research has focused on methods of displaying information to drivers in order to aid their attention to the road. This has involved a wide array of presentation methods (Capallera et al., 2022), including LED strip bars on the dashboard (Löcken et al., 2016; Yang et al., 2018), ambient lights around the steering wheel (Matviienko et al., 2016; Mok et al., 2017), or thermal and tactile cues around the steering wheel (Di Campli San Vito et al., 2018, 2020a). However, one of the most commonly suggested methods is to present information at the drivers' eye level with the road, known as a Heads-Up Display - (HUD) (Niu et al., 2024).

This is in contrast to traditional infotainment systems which display any non-driving activity on the centre console, known as a Heads-Down Display - (HDD), requiring the driver to take their eyes off the road and look down into the vehicle interior.

Static projection-based HUDs have been shown to be beneficial for presenting information to drivers, keeping their eyes on the road (Pauzie, 2015) and reducing reaction times towards road signs (Liu, 2003) and hazardous road events (Liu and Wen, 2004b). However, the type of information on display has typically been limited to simple informational displays, showing things like the current speed limit or basic navigational instructions (Pauzie, 2015). In recent years, there has been a drive for incorporating mixed reality displays into cars as infotainment systems (Basemark, 2024), in order to enhance productivity and entertainment during automated driving (McGill et al., 2017, 2022). Augmented Reality - (AR) in particular, where virtual content is overlaid on top of the real world (Azuma, 1997), can be used to provide a greater source of information over the projection HUDs that only display simple information. AR HUDs enable dynamic real time displays that can enhance the view of the road and provide information to drivers (Schroeter and Steinberger, 2016; Schömig et al., 2018) and have been shown to provide notable benefits to drivers in automated vehicles (Jing et al., 2022; Wu et al., 2023b).

However, there has been little research exploring the potential benefits of using an AR HUD for displaying NDRTs to drivers, making use of the dynamic nature of the interface to enhance the driver's awareness of the road while their attention is on the secondary task. Meta-reviews from Niu et al. (2024) and Riegler et al. (2019b) both highlight how there has been little research evaluating the potential for using AR HUDs to aid driver awareness of the road while they are engaged with an NDRT. While the evidence that displaying driving-related information at eye level in a HUD can be beneficial is mostly consistent (Liu and Wen, 2004a; Park and Park, 2019; Smith et al., 2023), preliminary research found that displaying an NDRT in an AR HUD impacts driver performance and response times to takeovers (Radlmayr et al., 2018; Hungund and Pradhan, 2023), eliminating the potential benefits of a heads-up view.

In order to allow drivers to safely engage with non-driving activity while they also remain responsible for supervising an automated vehicle, it is necessary to investigate the impact of interacting with NDRTs on the driver's ability to maintain attention and respond to danger on the road. If this is disrupted by an NDRT, is it possible to design an AR display which can aid driver awareness by cueing attention towards important aspects of the road scene even while they are performing an NDRT? Answering these questions will aid in the design of AR interfaces which allow drivers to engage with these NDRTs as they would like to without sacrificing the benefits of automated driving.

1.2 Thesis Statement

Semi-automated vehicles allow drivers to engage with non-driving related tasks (NDRTs). However, these tasks interfere with the driver's situational awareness, key to safely retaking control of the vehicle when required. This thesis investigates how Augmented Reality (AR) can be used for presenting NDRTs in order to reduce their impact on situational awareness. Over six experiments, the results show that simply presenting an NDRT using AR does not provide any inherent benefits. However, including attentional cues into the NDRT can aid situational awareness and even maintain it at the same level as without a distraction. This provides insights for the design of in-car interfaces that aid driver awareness during automated driving.

1.3 Thesis Research Questions

This thesis aims to address the following overarching research questions:

- **Research Question 1:** How does engaging with a distracting NDRT affect the ability to react to the road scene? (Chapter 3 - Experiments 1 and 2)
- **Research Question 2:** How does the location where a distracting NDRT is presented affect the ability to maintain awareness of the road scene? (Chapter 5 - Experiments 3 and 4)
- **Research Question 3:** How does the design of attentional cues affect ability to maintain awareness of the road scene during a distracting NDRT? (Chapter 6 - Experiments 5 and 6)

1.4 Thesis Structure

Chapter 2 aims to give a general introduction to the relevant academic literature and prevalent problems associated with the performing NDRTs during automated driving. Models of driver attention in both manual and automated vehicles are described, along with the challenges that come with presenting information to drivers of automated vehicles. Proposed solutions which attempt to overcome these challenges and the areas which this thesis addresses through the overall research questions are discussed.

Chapter 3 sets out the first two experiments, which employ a dual-task paradigm to investigate how performing a distracting secondary task presented in AR affects reaction time on a cognitive task. The first experiment used the Go/No-Go task, an inhibition control task where participants respond to Go signals and ignore No-Go signals.

Forty-two participants reaction times were measured both with and without performing a secondary AR task, either localised to the position of the signal or over a larger presentation area. This AR task involved fixating on virtual 3D gems appearing at random intervals to pop them and make them disappear. The second experiment followed a similar paradigm, but measured performance on the Hazard Perception test, a reaction time measure of driver performance. Thirty-six participants viewed 20 Hazard Perception video clips, 10 without a distraction and 10 while also engaged with the same distracting AR task as the first experiment. Once again, performance was compared between when the AR task was localised either to the centre of the video clip or over the whole screen. In line with previous research, both studies showed an overall increase in reaction time when participants were engaged with a distracting task in AR. However, there were no differences evident between the size of the presentation area for the AR tasks. The findings from these two experiments confirmed that performance is hindered when engaged with a secondary task, both on cognitive and driving-related reaction time tasks.

Chapter 4 describes the experimental methods used to evaluate the latter research questions of this thesis. An overview of the empirical methods used to assess driver attention and situational awareness in a psychological setting is provided. The creation of a Hazard Prediction test and subsequent validation study are described. This experimental paradigm was used for the studies in Chapters 5 and 6. This chapter provides an overview of the empirical methods used to evaluate driver attention and situational awareness throughout the rest of this thesis and serves as a reference chapter.

Chapter 5 describes two experiments which further investigate the effect of engaging with a distracting NDRT presented in AR on situational awareness, a more comprehensive measure of a driver's ability to perceive the road environment. The third experiment compared HUD and HDD presentations of a non-driving task displayed either via an AR headset or on a tablet computer. Thirty two participants performed the same eye-tracking distraction task as in Chapter 3 while responding to the hazard prediction clips described in Chapter 4. The results suggested that presenting the NDRT as a HUD did not provide any additional benefit, unless an attentional cue was included. However, all conditions saw poorer performance than when viewing the hazard clips without distraction. Following this, the fourth experiment measured hazard prediction performance while performing a a more realistic distracting task taken from the automotive literature: inputting mobile phone numbers onto a virtual keypad. An expert design focus group with 6 automotive user interface and human-computer interaction experts was conducted to design the attentional cue to be used for this task. Twenty four participants then viewed the same hazard clips presented in Chapter 4 while also performing this keypad task, either as a HUD or a HDD.

It was found that even with the attentional cue drawing attention to a specific part of the screen, hazard prediction performance was significantly impaired in all conditions compared to viewing the clips without distraction. Eye tracking data collected in AR indicated that the increased demand of the task meant attention was drawn to the cue itself rather than the driving task underneath. The findings from these experiments suggest that simply showing non-driving content at eye-level with the road does not confer the supposed benefits of a heads-up view. Similarly, the design of an attentional cue must be considered with the demand of the non-driving task.

Chapter 6 presents the final two experiments that form this thesis. These sought to investigate the effects of using social information to convey danger on situational awareness compared to a purely visual cue. In the fifth experiment, 24 participants performed the same AR task as in Experiment 3, while responding to the same clips as in Chapter 4. Hazard prediction performance was compared between a positional visual cue which moved to the location of the hazard, a virtual agent which oriented its head towards the location of the hazard, an NDRT only condition with no attentional cue, and a Control condition with no AR task. The results showed that performance was significantly worse than Control in all NDRT conditions, regardless of the inclusion of an attentional cue. Following on from this in the sixth experiment, the attentional cues were augmented to include an active hazard alert. The same NDRT and hazard prediction clips as Experiment 5 were used, and performance was compared between an active visual cue which change colour and moved to the location of the hazard, a virtual agent which oriented its head towards the location of the hazard and changed colour, a social agent which was animated to react to the danger and point towards the location, and a Control condition with no AR task. The results from this study found that there were no differences in hazard prediction performance between the colour-change cue and Control conditions, though for the animated social agent condition which had significantly lower scores than all of these conditions. The findings indicate that it is possible to use attentional cues to maintain driver awareness while they are performing an NDRT. Furthermore, it is possible to use social information to convey danger to similar effect as using a visual cue if used with a specific attentional alert to attract driver attention. However, the design of the social cue must be considered with regards to reducing distraction and ambiguity.

Chapter 7 provides an overall summary of the main findings from this thesis, in the context of the overall research questions posed. The contributions of these findings are discussed in an applied setting, as well as the psychological implications of presenting non-driving tasks in automated vehicles. The chapter concludes with a discussion of future research directions and reflection on the impact of the findings of this thesis.

Chapter 2

Literature Review

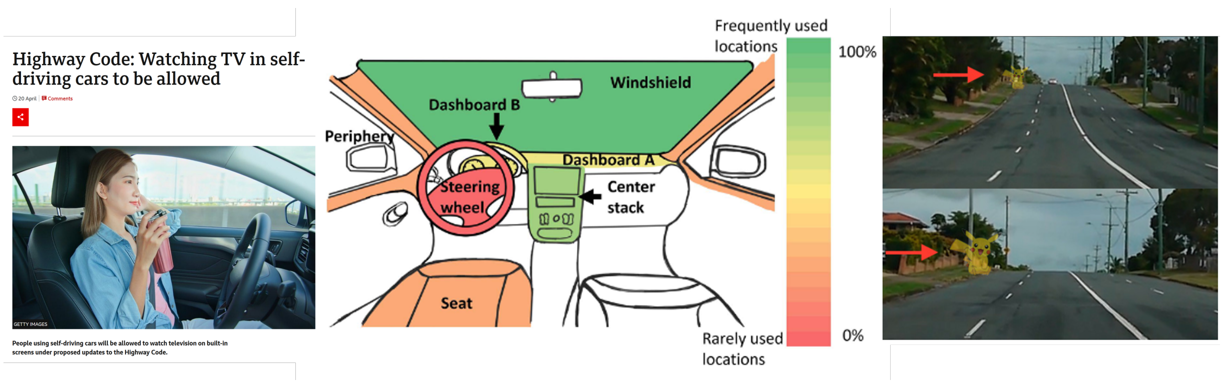


Figure 2.1: An article from BBC News (2022) describing changes to the Highway Code (left). A figure showing distribution of where driver displays are presented from Capallera et al. (2022) (centre). Concept for a gamified AR HUD to direct driver attention taken from Schroeter and Steinberger (2016) (right).

.

This chapter describes the literature that informs the overall thesis research questions set out in Section 1.3. Driver attention when manually controlling a vehicle has seen much research looking at how attention is maintained and distractions avoided to reduce the risk of collisions, outlined in Section 2.1. However, as control of the driving task is handed over to a computer and drivers are able to engage with non-driving activities, attention to the road in automated vehicles will have different requirements and constraints, which are described in Figure 2.1. Methods of displaying information about the road to the driver to aid their awareness of the road are set out in Section 2.2. However, it is important to evaluate the impact these displays have on a driver's ability to process information about the road, especially if distracted by non-driving activity. Therefore, Section 2.3. discusses the strengths and weaknesses for some of the proposed solutions for how overall driver awareness can be assisted during an automated drive.

2.1 Driver Attention

Understanding the Road

Driving is a complex cognitive task that requires the co-ordination of multiple physical and mental processes. Operationally, a driver needs to adjust the steering wheel to maintain their road position, operate the pedals to maintain an appropriate speed and interact with buttons and levers inside the car to signal their intentions to other drivers. Cognitively, they need to be aware of their current location, navigate towards their destination, as well as maintain awareness of the state of their vehicle. To do this safely, this awareness also needs to extend to other road users in the environment, their intended actions and how they may affect the driver's current choice of action. Many factors need to be taken into account to inform these decisions, which can only be done when a driver is fully aware of their road environment (Chaparro et al., 1999; Walker et al., 2009). Understanding these complicated interactions and processes has been the focus of much psychological research in order to highlight what the cognitive processes are that make a safer driver. This has been with the aim to provide training for novice drivers who have not yet have acquired these skills (Deery, 1999; Borowsky et al., 2010), but also to understand what has failed during these processes in the event of a collision (Yang et al., 2021b).

Gaze behaviour has been identified as an important factor when evaluating how a driver perceives the road. The ways in which novice and experienced drivers fixate on the road differs (Underwood et al., 2003) and changes as a driver becomes more experienced. Studies comparing the fixation patterns of experienced drivers found that they spend more time focusing on the road further ahead, compared to novice drivers who focus on the instrument cluster and the immediate road in front of them (Mourant and Rockwell, 1972; Underwood et al., 2003; Underwood, 2007). Additionally, experienced driver search patterns show more horizontal deviation towards areas of potential hazards, e.g., side roads or parked vehicles (Mourant and Rockwell, 1972; Underwood et al., 2003). This has the consequence that experienced drivers are greater able to anticipate future road events when they occur than novice drivers by picking up on anticipatory cues of what is about to happen (Crundall, 2016). For instance, a parked vehicle at the side of the road turning its wheels outward to change the direction of the car suggests that it is about to pull out into the road. Experienced drivers have been shown to notice these hazard precursor cues significantly more than novice drivers (Mourant and Rockwell, 1972; Lehtonen et al., 2014).

This ability to perceive anticipatory cues in a road environment has been reliably linked to involvement in collisions (Deery, 1999; Horswill and McKenna, 2004; Horswill et al., 2010), which are typically over-represented by novice drivers (Department for Transport, UK Government, 2023). As such, this *hazard perception* skill (Horswill, 2016) has been highlighted as an important ability of a safer driver. Evaluating this skill in the form of the Hazard Perception test (Horswill and McKenna, 2004) has been implemented as a requirement in the licensing procedures in multiple countries (Crundall et al., 2021; Driving and Vehicle Standards Agency, 2023). This test involves showing video clips of hazardous road events to would-be drivers, who respond as soon as they have noticed a dangerous event (Horswill et al., 2015; Driving and Vehicle Standards Agency, 2023), with their reaction time graded dependent on how soon they reacted to the onset of the hazard. More experienced drivers have been shown to reliably outperform novice drivers in this style of test (Sagberg and Bjørnskau, 2006; Borowsky et al., 2010), perceiving dangerous road events sooner and reacting faster than their novice counterparts.

Looking but Failing to See

However, perceiving a hazardous event is not the same as cognitively appraising and understanding it. While the hazard perception test is useful for differentiating between drivers in a test setting, it is not representative of the more complex processes a driver exhibits when out on the road (Crundall, 2016). Many studies demonstrate the concept of *inattentional blindness*, (Simons, 2000) where humans fail to adequately process information presented to them (Hills, 1980; Wolfe et al., 2022). This was famously demonstrated by Simons and Chabris (1999), who found that participants asked to count the number of passes players in white shirts made in a video of basketball, did not notice a person dressed in a gorilla costume walking across the screen. This finding has been replicated in a multitude of different scenarios, including missing information on advertisement banners (Resnick and Albert, 2014; Gelderblom and Menge, 2018), expert radiologists missing a gorilla image superimposed on a lung CT scan (Drew et al., 2013), and participants not noticing a sudden image of a spider appearing during a visual discrimination task (Wiemer et al., 2013). Subsequent studies which tracked the gaze patterns of participants in these types of experiments highlighted that, even when participants were fixating on the "gorilla", they would not process it if they were distracted by another task (Pappas et al., 2005; Memmert, 2006; Gelderblom and Menge, 2018; Drew et al., 2013). This is evident when the perceptual load of a task increases, meaning that the more complex a task, the greater the risk of inattentional blindness (Matias et al., 2022). This effect has also been shown to exist in a driving context (Brown, 2002). Experienced drivers fail to accurately recognise bicycles (Summala et al., 1996), motorcycles (Crundall et al., 2008) and police cars (Langham et al., 2002) at T-junctions, despite actually fixating on them.

This has been reasoned to be because drivers rely on their previous experience of the road and anticipate more commonly occurring dangers to them, thus failing to properly process the actual road scene in front of them in favour of what they are expecting to see (Summala et al., 1996; Crundall et al., 2008). Motorcycle riders who also drive a car do not make these same errors in attention (Magazzù et al., 2006) as they are more familiar with the idea of a motorcycle on the road. Similarly, drivers who are also cyclists are more effective at scanning for vulnerable road users (Kaya et al., 2021). While this has been shown to be dependent on the prevalence of road users (Beanland et al., 2014), more common objects in the road scene are processed faster and more readily. The existence of this phenomenon indicates that perception alone does not fully account for the ways in which drivers cognitively process the road scene.

Situational Awareness

Key to navigating a complex road scene is to be cognizant of all other road users and features. However, in order to make appropriate driving decisions, it is arguably more important to be able to anticipate what might happen next in the scene. For example, when driving down a busy high street, a driver must: a) be aware of the behaviour of vehicles ahead to ensure they do not suddenly stop, b) be mindful of any side roads or parking spaces to see if there is a car that may be about to pull out, and c) anticipate the behaviour of a pedestrian who might be about to step into the road after turning their head to check if it is clear. While on paper this is a highly complicated cognitive task, experienced drivers are able to effectively monitor all of these potential hazards and anticipate their actions in order to navigate safely through this environment.

Successfully being able to process this type of complex environment has been described by Endsley (1995b) as possessing a state of Situational Awareness - (SA). Endsley (1995a) proposes that there are three stages to this: 1) *Perception* – being able to perceive what is currently in the environment, e.g., a driver notices the wheel of a parked car turn outward, 2) *Comprehension* – understanding the causes behind the occurrence of an event and in the environment, e.g., realising that a turning wheel indicates that the driver of the parked car is preparing to pull out and join the road. 3) *Projection* – anticipating the future possibilities of what could happen, e.g., predicting that this parked car will start to move off into the path ahead (see Figure 2.2). While perception of the road scene is important, this final *Projection* stage draws upon prior experience of similar situations in memory to inform the decisions that a driver takes (Endsley et al., 1998).

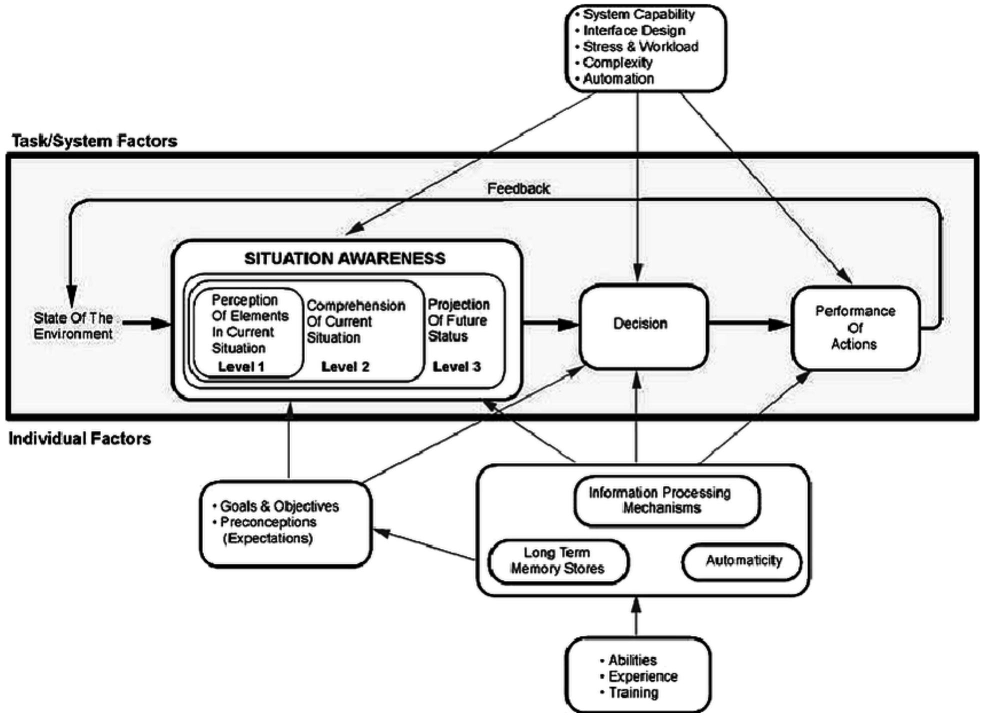


Figure 2.2: Endsley (1995)'s model of Situational Awareness.

Driving down the same busy high street on multiple occasions, a driver will encounter the above example frequently, and so will be primed to notice the anticipatory clue of a wheel turning out and accurately predict when a parked car intends to pull out. They can then take appropriate action to navigate the road safely, such as slowing down to allow the car to pull out in plenty of space. As stated previously, experienced drivers are more effective at noticing these anticipatory cues that inform these predictions (Mourant and Rockwell, 1972; Lehtonen et al., 2014; Crundall et al., 2017). There is also an element of automaticity, with highly experienced drivers reacting automatically to road events they are familiar compared to novel encounters (Charlton and Starkey, 2011).

However, this can lead to visual attention shifting towards more irrelevant stimuli on these familiar routes (Young et al., 2018; Harms et al., 2021), which contributes to increased inattentional blindness to novel events (Charlton and Starkey, 2013). This manifests in the highest rate of collision occurring within a 10-minute journey from home (Chen et al., 2005). However, this is only apparent in experienced drivers, and not novices (Burdett et al., 2017). This demonstrates the importance of remaining situationally aware, regardless of driving ability or amount of driving experience, throughout the whole drive to be able to anticipate hazards on the road.

Attention in Automated Vehicles

The majority of the research into driver attention has been focused on manual driving, where the driver is fully in control of vehicle. It emphasises the importance of a driver being situationally aware as they are the ones that need to make the appropriate decisions to safely operate the vehicle. However, this is starting to become less important with the increase in the availability of Automated Vehicles - (AVs), where parts or all of the driving task can be outsourced to a computer. The tasks that automated vehicles can perform has significantly expanded in recent years to controlling the steering, keeping within the lane markings on the road, and matching the speed of the vehicle ahead (Mallozzi et al., 2019). More recent advances even allow navigation of a long-distance journey with little intervention from the driver. These different automated functions have been categorised by the Society of Automotive Engineers (2018) into 5 levels of automated driving, which set out how an AV can be classified depending on the tasks it can perform, ranging from manual control of the vehicle to full automation (see Table 2.1). As the sophistication of AVs increases, there is less and less need for a human to involve themselves with first the physical, then the cognitive demands of driving. The result of this is that the cognitive relationship with the driving task will change from an attentional task to a supervisory one.

SAE Level		Description	
Level 0	Full Manual Control	The driver has full responsibility for the vehicle and the driving task	Human operated
Level 1	Driver Assistance	The vehicle performs limited driving tasks, e.g., cruise control	Feet off
Level 2	Partial Automation	The vehicle performs multiple driving tasks, e.g., steering, acceleration control	Hands off
Level 3	Conditional Automation	The vehicle performs the driving task but a human driver is required to supervise	Eyes on
Level 4	High Automation	The vehicle is responsible for the majority of the driving task	Eyes off
Level 5	Full Automation	The vehicle is responsible for the whole driving task and there is no need for a human present	Mind off

Table 2.1: A table listing the 5 levels of automated driving as set out by the Society of Automotive Engineers (2018).

Unfortunately, despite the promise of fully replacing drivers with computers, the progression of automated vehicle technology has not matched the enthusiasm for the benefits it can provide (Hancock, 2019). Vehicles at higher levels of automation are currently confined to specific geo-fenced areas, such as in the US for Waymo and Uber (Axios, 2023; TechCrunch, 2024). As such, AVs currently available to consumers still require them to maintain vigilance and continue paying attention to the vehicle while the driving task is automated. This is starting to become a legal requirement in many countries (UK and Scottish Law Commissions, 2020; The European Parliament, 2022; UK Government, 2022b), where the drivers are held responsible for maintaining attention to the road throughout an automated drive. This means that, for the foreseeable future, drivers will be able to engage in automated driving but are still required to maintain their attention to the driving task and the vehicle to ensure its smooth operation. Subsequently, the way in which drivers process the road will change.

Gaze patterns have been shown to differ between manual and automated driving shifting towards anticipation (Mars and Navarro, 2012; Mole et al., 2021), with fixations located closer to the front of the vehicle during an active drive and greater horizontal scanning in a passive drive (Mackenzie and Harris, 2015). This change in responsibility also means that the actual role of the driver becomes less of an operator and more of a passenger (McGill et al., 2017). They will spend more time supervising the operation of the vehicle than manually controlling it. Unfortunately, humans are not cognitively suited for supervising automated tasks (Bainbridge, 1983), with attention and performance of the supervision task decreasing as time spent supervising continues. This is evident when a driver supervises an AV, with low mental workload during automated drives resulting in reduced awareness of the driving task (Bourrelly et al., 2019). Driving performance, such as reaction time to takeover requests, significantly decreases as time spent driving increases (Thiffault and Bergeron, 2003; Ting et al., 2008), and is also associated with feelings of boredom (Schroeter et al., 2014) and fatigue (Lal and Craig, 2001; Figalová et al., 2023). Notably, drivers respond significantly slower to safety critical events during automated driving (Merat and Jamson, 2009; Louw et al., 2015).

If drivers struggle to maintain SA whilst fully in control of a manual vehicle, this is only worsened when they are supervising an AV where there is no requirement or motivation to be engaged in the driving task (Endsley, 1995a; Pipkorn et al., 2022). The challenge therefore is, while AVs still require a human to be ready to resume control of the vehicle, how to keep drivers situationally aware of the current driving environment during an automated drive.

2.2 Displaying Information to Drivers

Human-Machine Interfaces

Designing systems to provide SA information to drivers is an important step in the development of autonomous systems (Endsley, 2017). As such, a wide array of interfaces for displaying information to drivers have been developed. Currently, these are limited to displays of driving-related information, such as current speed or navigational aids. The goal of these is ultimately to provide information to the driver about the state of the vehicle, known as Human-Machine Interfaces - (HMIs). In a driving context, the goal of a HMI is to communicate information to the driver in a way that is easily interpretable and does not distract from the driving task. However, the role of an HMI for communicating information to a driver becomes much more important in an AV where they are not necessarily attending to the road. The current system to indicate when the driver should take over control of the vehicle is to use an attentional alert; a Takeover Request - (TOR).

These are designed to capture attention and inform a driver when they should resume responsibility for the driving task, either through a planned handover or an emergency interjection. A significant amount of work has been invested in researching and designing TORs that can alert the driver in the most effective and informative ways (Oviatt, 2006; Peck et al., 2015; Salubre and Nathan-Roberts, 2021). The challenge with using a HMI to convey information visually is that driving is predominantly a visual task. Most of the visual cues drivers use to anticipate the road scene come from things the driver can see (Crundall, 2016). As such, there is a cognitive cost for presenting information in a way that encroaches on processing of the driving task. According to Wickens (2002)'s Multiple Resource Theory, cognitive performance is impaired when two tasks whose domains overlap are performed. For example, asking participants to perform two auditory or two visual tasks simultaneously leads to a reduction in performance on both tasks (Treisman and Davies, 2012). This finding extends into the driving domain, where attending to visual displays inhibits hazard response times (Horrey and Wickens, 2003) and draws attentional resources away from the driving task, impairing driving performance (Horrey et al., 2006).

Therefore, designing an HMI that draws attention away from the road is counterproductive if the goal is to enhance a driver's awareness of the road. Wickens (2002) describes how splitting visual task between central and peripheral fields of view has reduced impact on dual-task performance. This has been applied in the aviation domain, with peripheral HMI displays aiding fighter pilots in a way that reduces the impact on the main attentional task (Prinzel III and Risser, 2004). This has been used to inform the design of in-car HMIs to create displays with less impact on attention. A review by Capallera et al. (2022) identifies how the majority of interfaces for supporting SA focus on peripheral interactions, such as including LED strip bars on the dashboard (Yang et al., 2018; Löcken et al., 2016), ambient lights around the steering wheel (Mok et al., 2017; Matviienko et al., 2016), combined dashboard and centre console displays (Wang et al., 2017a) and peripheral lights on a pair of glasses (Van Veen et al., 2017). Outside these typical modalities, the use of haptic vibration and thermals cues has been suggested for portraying information to drivers (Di Campli San Vito et al., 2020a), and has been shown to aid with navigation tasks (Di Campli San Vito et al., 2017), lane change manoeuvres (Di Campli San Vito et al., 2018) and transfer of control between the driver and the AV (Di Campli San Vito et al., 2017).

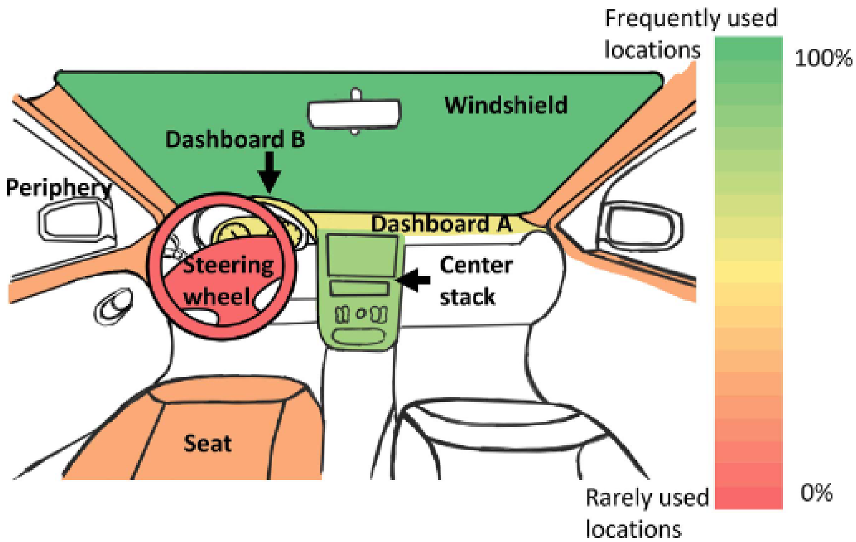


Figure 2.3: Image showing the frequency of interaction location taken from Capallera et al. (2022)'s literature review.

An alternative approach is to design an HMI that captures the attention of the driver quickly and directly. Multimodal interfaces, where two or more different sensory types of alerts are combined, e.g., auditory and tactile, have been shown to be more effective at alerting drivers to danger than unimodal interfaces (Politis et al., 2013). Using multimodal cues as driver alerts increased the perceived sense of urgency and reduced the reaction time to a TOR (Politis et al., 2013, 2014b) over unimodal alerts in both normal and safety-critical handovers (Politis et al., 2017), a common finding when evaluating multimodal vehicle HMIs (Mehrotra et al., 2022). However, including multimodal feedback without a visual component can be seen as disruptive (Di Campli San Vito et al., 2017), and using more modalities for conveying urgency can be seen as annoying (Politis et al., 2014b). Yang et al. (2022) highlight how the use of specific modalities impacts the effectiveness of a multimodal TOR, and it is not as simple as *"more is better"*.

Presenting Non-Driving Information

This raises an issue concerning the importance of including visual components of a multimodal interfaces is that driving is a visually demanding task. An additional problem is that drivers report that their motivation for buying an AV is not to watch the car drive itself (Merfeld et al., 2019). Activities such as socialising with passengers, reading, being productive and entertainment via watching a film or using a personal device (Panagiotopoulos and Dimitrakopoulos, 2018; Merfeld et al., 2019; Wilson et al., 2022) are described as advantages of automation, which contain a visual element. Despite the benefits of these HMIs described previously, their main benefit is to convey information to a driver who is attentive to the road and facilitate them taking back control.

Furthermore, the majority of these studies measure the effectiveness of each HMI when the driver is paying attention to the driving task. What is not clear is whether these HMIs can benefit a distracted driver who is engaged with a Non-Driving Related Task - (NDRT), something which is not relevant to the operation or supervision of the vehicle.

Impact of Multitasking on Attention

In manually driven cars, distraction from NDRTs has been shown to be a major influence on road safety. UK Government road safety statistics indicate that 34 % of collisions in 2022 were caused by distracted or impaired driving (UK Government Statistics, 2024). The most common source of this distraction is a mobile phone, which accounts for approximately 8% of fatal collisions in the US (Forbes Advisor, 2024). Research investigating the use of mobile phones while driving demonstrates the cognitive deficits caused by interaction with a distracting secondary task.

Mobile phone use has been directly linked to collision risk (Ige et al., 2016; Lipovac et al., 2017), and, as such, using a mobile phone has been outlawed while operating a vehicle (UK Government, 2022b). The deficit introduced by engaging with a distracting task has been explored extensively in research investigating dual-task attention. This is a paradigm where multiple tasks are carried out concurrently, and performance compared to when performing a single task (Heuer, 1996). The dual-task paradigm has shown that performance deteriorates for a variety of cognitive tasks (Pashler, 1994; Pashler and Johnston, 1998; Fischer and Janczyk, 2022) due to a bottlenecking of cognitive resources (Boles et al., 2007), though there is evidence this can be mitigated through practising the two tasks simultaneously (Strobach, 2020).

This problem of overlapping tasks is the main issue behind driver distraction when performing NDRTs while controlling a vehicle. Predictably, the attention of drivers in automated vehicles also suffers if they are engaged with a NDRT. Studies measuring driving performance after resuming control of an automated vehicle show a deterioration after drivers are distracted by secondary tasks (Merat et al., 2012; Pipkorn et al., 2022). While some studies suggest that NDRTs can reduce fatigue and increase attention during extended automation (Jarosch et al., 2017; McKerral et al., 2023), a meta-analysis of studies investigating take-over time identified the undertaking of visual NDRTs as a key factor that increases takeover time (Zhang et al., 2019). Furthermore, since it takes time for awareness of the road to be acquired following a TOR (Gold et al., 2013, 2016) this will be further impaired if a driver is distracted by an NDRT (Wandtner et al., 2018; Dogan et al., 2019).

Despite this, consumers report wanting to engage with NDRTs including socialising, leisure activities and productivity tasks (Panagiotopoulos and Dimitrakopoulos, 2018), not having to remain focused on the road during an automated drive. This lack of motivation to maintain vigilance during an automated drive, paired with displays which may not provide adequate situational awareness, has dangerous implications as AVs become more commonplace on the road; particularly if drivers do not fully understand the limitations of automated driving technology (Dixon, 2020). Consequently, it is important to consider how HMIs can be designed and used to provide information to a driver who is distracted by a NDRT.

2.3 Enhancing Driver Awareness

A common attribute of these distracting NDRTs is that they take the driver's gaze, a precursor for attention, away from the road ahead and towards driving-irrelevant features. Non-driving tasks are typically presented in the centre stack as a Heads-Down Display - (HDD). Tasks such as changing the radio or inputting navigation information require a driver to take their eyes off the road to interact with them, impairing their ability to attend to the road. The same is true for a driver in an AV who wants interact with a personal device or perform an NDRT on the centre console. To counteract this, presenting information at eye-level with the road as a Heads-Up Display - (HUD) has been suggested to reduce this eyes-off-road time. Providing information to drivers in a HUD has seen benefits to steering behaviour (Liu, 2003; Liu and Wen, 2004b), as well as increasing the duration of fixations on the road ahead, compared to without a HUD (Liu, 2003; Liu and Wen, 2004b; Zhang et al., 2021). Furthermore, reaction to hazardous events (Horrey et al., 2003; Liu and Wen, 2004b) and driving performance (Smith et al., 2023, 2016) was superior when engaging with a HUD compared to a HDD. As shown in Capallera et al. (2022)'s review, the majority of interaction locations focus on the windshield, to take advantage of the fact the driver is looking ahead at the road (see Figure 2.3). It is sensible to suggest then that presenting a NDRT as a HUD may also aid driver attention.

Using Augmented Reality to Cue Attention

One of the most commonly proposed applications of a HUD display is for cueing driver attention (Yeh and Wickens, 1999; Riegler et al., 2019b). To do this requires a more sophisticated interface than one which just displays static information. Using an Augmented Reality - (AR) display, where holographic content is overlaid onto the real world view (Azuma, 1997) allows for a mixed reality HUD where the information displayed can dynamically change and react to events that occur in the real road environment (Gabbard et al., 2018; Schömig et al., 2018).

While traditional HUDs typically consist of a static projection of information, AR HUDS can enhance what can normally be perceived by supplying additional information than what is typically available, e.g., dynamic navigational information that updates as the vehicle moves or highlighting potentially important road features in real time (see Kettle and Lee (2022) for a review). Previous work looking at utilising in-car AR HUDS suggests that they can also aid driver attention on the road and driving performance over a HDD (Medenica et al., 2011; Langlois and Soualmi, 2016). Jing et al. (2022) found that AR HUDs were able to reduce distraction when focusing on dangerous driving scenarios, and Bark et al. (2014) showed that a navigational AR HUD aided turn decisions, though this differed between 2D and 3D displays. Additionally, Lindemann et al. (2018) found that AR HUDs showing a variety of driving-related information such as threat markers and oncoming traffic indicators improved situational awareness of drivers.

In an AV context, de Oliveira Faria et al. (2021) found that AR cues helped improve driver behaviour after a TOR as well as reducing the number of driver-initiated TORs. Furthermore, the use of AR HUDs as a driver awareness aid, e.g., a crash warning system, has been shown to help reduce mental workload (Haas and Van Erp, 2014; Bauerfeind et al., 2021), reduce reaction times (Kim et al., 2013), reduce gaze towards the instrument cluster (Schömig et al., 2018), and reduce boredom (Schroeter et al., 2014; Steinberger et al., 2017). This suggests that using AR can also be beneficial for facilitating SA during automated driving. Specifically, cueing attention using an AR HUD has been shown to benefit target detection (Yeh and Wickens, 1999), visual search at a distance (Warden et al., 2022), and a virtual-navigation task (Stefanucci et al., 2022). Karatas et al. (2020) showed that using an AR HUD highlighting hazards led to quicker recognition compared to a traditional HUD (Karatas et al., 2020). Similar findings have been demonstrated that specifically directing attention with AR cues increased detection rates of pedestrians and road targets (Rusch et al., 2013; Wang et al., 2017b; Kim and Gabbard, 2022; Wang et al., 2022b).

Presenting NDRTs in AR?

While these studies demonstrate how an AR HUD presentation benefits driver attention when attending to the road, the effect of presenting a NDRT via an AR HUD is not clear. There have been suggestions to *gamify* the supervision task and add game-like elements that also provide situational awareness information (Schroeter et al., 2014). Schroeter and Steinberger (2016) describe an AR HUD concept where game content overlaid on the driving scene can encourage drivers to maintain attention on the road during automated driving (see Figure 2.4). It has been suggested that this would aid drivers awareness of the road scene in the event of a TOR.

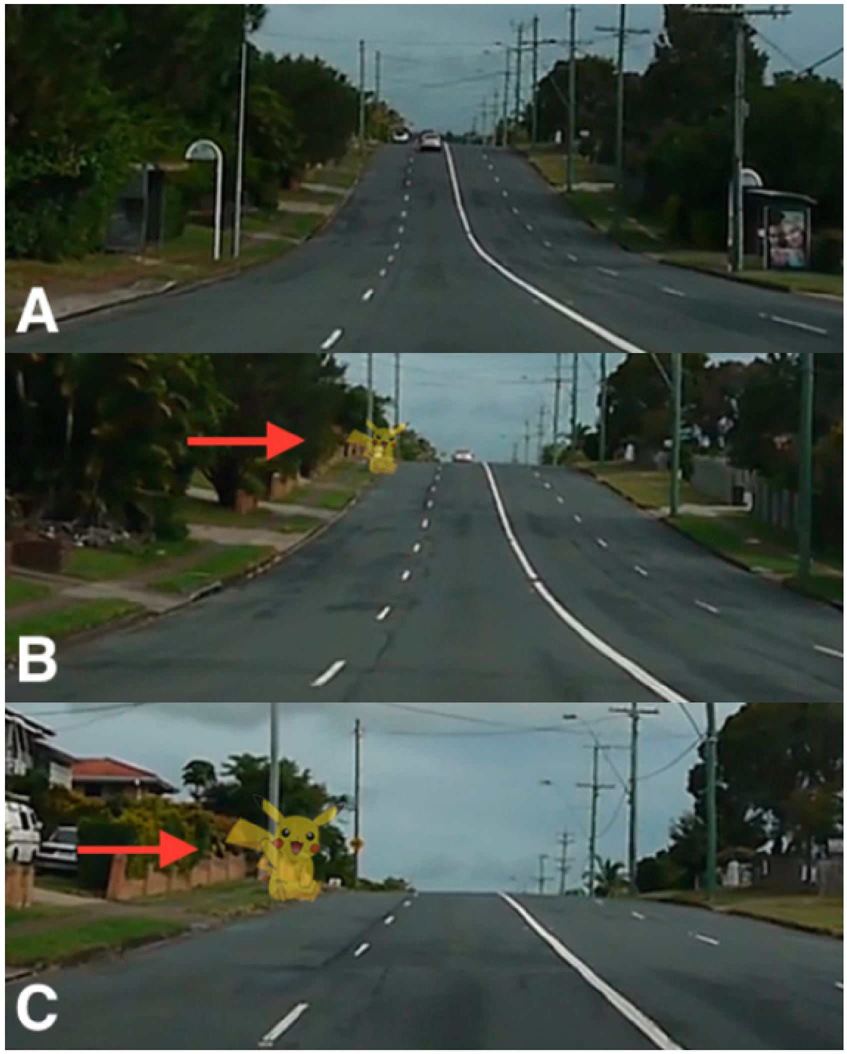


Figure 2.4: The Pokemon Drive AR concept, which shows gamified AR elements to encourage attention to hazards, taken from Schroeter and Steinberger (2016) (Note, the red arrow is for the reader only and not part of the concept.

Wu et al. (2023a) also proposed a SA-based in-car game which uses external road features and objects as game elements, similar to the in-car zombie shooting game for passengers proposed by Togwell et al. (2022). They report higher overall subjective SA levels while playing the SA game compared to a video. However, a study from Radlmayr et al. (2018) which used AR to present NDRTs demonstrated how a distracting balloon popping game for drivers to play peripherally significantly impaired driving performance, to a similar level as driving in thick fog. This research would suggest that, while AR HUDs are beneficial for enhancing driver awareness, their benefits when displaying a NDRT are not as clear.

Drivers in automated vehicle may be able to take advantage of the presentation at eye level with the road, but their ability to actually process the road scene has not been yet fully evaluated. There is a risk that, due to the cognitive limits caused by multitasking described in section 2.1, the distracting nature of the NDRT means that the benefits of a HUD display are lost. It is not clear whether the use of an AR HUD specifically for displaying NDRTs serves as an aid or a distractor for a driver's attentional resources.

Generally, studies investigating the potential benefits of HMIs or HUDs measure the effect of driver performance when supervising the driving task in a partially automated vehicle when their attention is solely on the road. While these concepts and studies indicate a benefit of AR HUDS for providing SA information, the research does not address the cognitive distinction between attention and awareness as previously discussed in Section 2.1. SA is likely to be differently affected if the driver is also engaged in a distracting NDRT.

A review by Riegler et al. (2019b) into the use of AR applications for automated driving found that most research focuses on the use of AR for safety and driver assistance, not for presenting NDRTs. Similarly, a meta-review by Niu et al. (2024) evaluating the use of HUDs to display warnings to drivers noted that, while they have benefits over HDDs, few studies investigate how using an AR HUD to displays these warnings is affected when it is showing non-driving related information.

In light of the psychological literature, there is a danger that presenting NDRTs and warnings in this way could cause inattentional blindness, where a distracted driver focusing on an AR HUD fails to process a hazardous event, despite looking towards the danger. To fully take advantage of the benefits that automated driving can offer, it is important to investigate whether presenting driving-related information as an AR HUD can provide the same benefits to attention and situational awareness when presenting a NDRT. If not, it is necessary to investigate whether it is possible to design an AR HUD which can provide this information to a driver while they are distracted by an NDRT.

2.4 Chapter 2 Summary

With an increasing prevalence of partially automated vehicles, the way drivers pay attention to the road is changing. Yet, since the development of more sophisticated automated vehicles has not occurred as quickly as previously thought, drivers of AVs are required to maintain attention to the road in case they need to regain control. This presents an attentional conflict between a distracting NDRT and the need to supervise the automated vehicle. There is a push towards incorporating mixed-reality interfaces as in-vehicle displays (Basemark, 2024), with presenting NDRTs being one of the described benefits of these (McGill and Brewster, 2019; McGill et al., 2022; Togwell et al., 2022). However, engaging with a distracting non-driving task has detrimental effects on a driver's ability to maintain awareness of the road, and impacts their ability to safely regain control of an automated vehicle. The effects of performing a distracting NDRT on the SA of a driver supervising an AV have not had the research attention that HMIs and driver displays have typically seen. Furthermore, the higher cognitive processing of SA and the cognitive implications of simultaneously attending to the road and a NDRT have not been explored.

Therefore, it is relevant to understand the cognitive underpinnings of how drivers would interact with NDRTs presented via AR HUDS, and the impact they could have on SA. Does a HUD presentation inherently benefit SA due to reducing the eyes-off-road time when a driver performs a NDRT? Or conversely, as research into multitasking and inattentional blindness suggests, does the distraction posed by a NDRT remove the benefits of a HUD presentation due to attentional constraints? Is it possible to design an AR HUD to present a NDRT that also provides SA information to the distracted driver?

This thesis attempts to further investigate these questions by measuring the awareness of drivers engaged with a distracting non-driving task presented in AR. This is structured around the 3 research questions that were set out in section 1.3. Chapter 3 will address the first research question:

- **RQ1) How does engaging with a distracting NDRT affect the ability to react to the road scene?**

This chapter sets out to establish what the impact of performing an NDRT on driver cognition is. This was done by comparing performance on measures of reaction time both with and without a secondary distracting NDRT task presented in AR. This was in line with trends in previous research that evaluate how quickly a driver can respond to events, such as a takeover request or a danger on the road. Following this, Chapter 5 will address the second research question:

- **RQ2) How does the presentation method of an AR interface affect the ability to maintain awareness of the road scene?**

This was approached through an evaluation of different presentation methods for an AR interface displaying an NDRT. An AR HUD that is presented at eye-level is compared to a HDD requiring eyes off the road, and the ability to predict future events rather than perceive current ones is empirically measured, using methods described in Chapter 4. Eye tracking data are also presented, which indicates that fixation on the driving task may not necessarily correspond to attention when a driver is distracted by a NDRT. Finally, Chapter 6 will address the third research question:

- **RQ3) How does the design of an AR interface and the inclusion of attentional cues impact the ability to process a road scene?**

Novel methods and modalities for displaying information to a distracted driver are explored with regards to efficiently processing information. With recent suggestions to include virtual assistants into the vehicle and frame interaction in a social capacity, the use of a virtual agent for presenting information to distracted drivers to aid their SA could provide benefits. A comparison of positional cues and hazard warning cues is presented between visual and social domains, which suggest that using social behaviour to provide cues to drivers can be as effective as a visual cue and maintain SA at the same level was when viewing the road without distraction.

The experimental methods that were used to approach these research questions and the results that were found are set out in the 4 subsequent chapters, with a discussion of the results for each experiment included at the end of each chapter. These results are summarised and discussed in the wider research context in the final chapter.

Chapter 3

Investigating the Effect of Performing a Distracting AR task on Reaction Time

Figure 3.1: Examples of the Augmented Reality cues from Experiment 1 (left - Go No-Go task with game presented in AR in front of road environment) and Experiment 2 (right - keypad task).

In this chapter, two experiments are presented which measured reaction time performance on an inhibition control task (Experiment 1) and a hazard perception task (Experiment 2) to address Research Question 1. Performance was compared between a baseline and a dual-task condition, where participants also performed a secondary task presented in AR. This task was split into two display sizes: Central (items presented in a smaller visual area) and Global (items presented over the whole AR field of view). The results from both experiments showed a significant increase in reaction time in the dual task conditions compared to the baseline single task conditions, addressing **RQ1** of the thesis. The implications of how these findings can inform the use of AR to present NDRTs is discussed, with regards to how driver attention is taken away from the road.

3.1 General Introduction

Reacting to the Road

Driving is a complex cognitive task which requires a driver to make quick appraisals when evaluating the road scene and then take safe and appropriate actions. Many factors must be taken into account to inform these decisions, which can only be done when drivers are aware and paying attention to the driving environment (Walker et al., 2009). Typically a driver is required to maintain full attention to the road in order to accurately scan for potential dangers and react to them accordingly. The ability to do this is known as their Hazard Perception skill (Horswill and McKenna, 2004) and is measured as a push-button response to a hazardous driving clip, with faster reactions scoring higher. This skill has been reliably linked with the likelihood of being involved with a collision (Horswill et al., 2010) with drivers that perform better at Hazard Perception exhibiting fewer crashes (Deery, 1999; Horswill et al., 2010). Engaging with a distracting secondary task, such as using a mobile phone, has detrimental effects on hazard perception, leading to slower reaction time (Törnros and Bolling, 2005; Haque and Washington, 2013; Yannis et al., 2014) and increased risk of collisions (Asbridge et al., 2013).

The advent of Automated Vehicles - (AVs) allows and encourages drivers to hand over control of the driving task to the vehicle. This allows drivers to engage with Non-Driving Related Tasks - (NDRTs) safely while the vehicle is responsible for perceiving and reacting to hazards. However, AVs currently available to consumers still require the driver to maintain supervision of the vehicle as a fail-safe (UK and Scottish Law Commissions, 2020). If an issue arises and automation fails, the driver needs to be ready to resume control of the vehicle and navigate the potentially dangerous situation at a moments notice. This presents a cognitive challenge to drivers in partially automated vehicles who want to engage with NDRTs. As in manual cars, engaging with a distracting secondary task in AVs has been shown to significantly increase reaction time to takeover requests (Louw et al., 2015; Dogan et al., 2019). The more demanding an NDRT is of a driver, the longer it takes for them to take control (Yoon and Ji, 2019) and when they do, it is usually in a less able state to deal with the road scene (Radlmayr et al., 2014). This is consistent with psychological literature on dual-task performance, where performance on two tasks performed simultaneously is poorer (Heuer, 1996; Schumacher et al., 2001). It poses the question of whether it is possible to display NDRTs to drivers in a way which limits their impact on attention.

Giving Drivers a Heads up (Display)

Previous work looking at Heads-Up Displays - (HUDs) suggests that driving performance was less impaired and preferred to traditional cockpit displays (Medenica et al., 2011; Smith et al., 2015; Jose et al., 2016). This has already become available in certain vehicles available to consumers (Firth, 2019b). The use of HUDs as a driver awareness aid, e.g., a crash warning system, has been shown to help reduce mental workload (Schömig et al., 2018) and reduce reaction times in manual driving (Kim et al., 2013; Wintersberger et al., 2018b). Yet, other studies have produced mixed results, suggesting that the design of the HUD has an impact on driver awareness (Kim and Gabbard, 2018, 2022). In an AV context, HUDs have been shown to reduce driver distraction when focusing on dangerous driving scenarios (Jing et al., 2022). In particular, using Augmented Reality - (AR) to show more dynamic displays benefits takeover times (Jing et al., 2022) and reaction times (Karatas et al., 2020). However, there are important considerations to be made for the design of HUDs, such as the demands of the task and how this affects gaze behaviour. The visual complexity of the driving scene and mental workload have been shown to affect a driver's eye movements (Crundall and Underwood, 1998; Chapman and Underwood, 1998; Kapitaniak et al., 2015), which in turn affects the detection of hazards (Mackenzie and Harris, 2015; Yang et al., 2021a). Similarly, a busy and attention capturing HUD is more likely to distract attention (Lee et al., 2020; Kim and Gabbard, 2022) and increase reaction times to dangers on the road (Wolffsohn et al., 1998). If a driver is engaged with a visually demanding display they have less attentional capacity available to focus on the driving task (Wickens, 2002).

Regarding eye movements, it has been shown that drivers tend to fixate on a central point in the distance where visual motion emanates from (Mourant and Rockwell, 1972), known as the focus of expansion. From here, experienced drivers deviate in a horizontal pattern to scan for potential hazards (Underwood et al., 2003), whereas novice drivers show a narrower spread of fixations clustered in the centre (Mourant and Rockwell, 1972; Crundall and Underwood, 1998). This leads to slower reaction times to developing hazards in novices (Horswill and McKenna, 2004), due to these poorer search strategies. Presenting information to drivers at eye-level has been shown to benefit attention, as it is less disrupting to these search strategies (Karatas et al., 2020; Wu et al., 2021) compared to looking down into the vehicle. Therefore, presenting a NDRT in this manner should show similar benefits, as it would allow the driver to interact with non-driving related content on top of the view out of the front of the vehicle while being able to monitor the road and to react to any hazards (Riegler et al., 2019b). However, the specific effects on reaction time of presenting NDRTs via a HUD is still not clear, in particular how an NDRT disrupting the typical eye movement patterns of drivers affects their reaction time.

On top of the questions raised over the benefits of AR HUDs, there are also practical issues to be considered. In particular, the possibility of motion sickness caused by the mismatch between an AR display and the dynamic driving scene behind it (McGill et al., 2017). There has been a wealth of research into simulator sickness in Virtual Reality - (VR) (Saredakis et al., 2020; Rangelova and Andre, 2018), and also looking specifically at in-car VR (McGill et al., 2017; McGill and Brewster, 2019; Pöhlmann et al., 2023). On the other hand, there is less research looking at sickness in AR interfaces. What has been done suggests that extended exposure to AR can also lead to sickness symptoms (Hughes et al., 2020; Kaufeld et al., 2022), and is comparable to VR induced sickness (Pettijohn et al., 2020). This is significant as placement of virtual content will have an important effect on how drivers are able to pay attention to the road. Häuslschmid et al. (2018) illustrated how, despite obscuring safety critical areas of the road, users still chose to place virtual content so it obscured these critical areas in the centre of their view. This is likely to lead to increased motion sickness due to the conflict between the static AR display and the dynamic driving scene (Duh et al., 2004; Kaufeld et al., 2022). Should in-vehicle AR become the presentation method of choice in AVs, it will need the same research focus as in-car VR to inform its design and application in order to ensure driver attention is not impaired and mitigate the onset of sickness symptoms.

This chapter investigates performance on reaction time tasks whilst simultaneously performing a distracting secondary task in AR. Two experiments are presented which compare single task to dual-task performance on an inhibition control (Experiment 1) and Hazard Perception (Experiment 2) task, to evaluate the impact of using AR display for an NDRT on attention. The effect of interacting with an AR-interface on a driving-related task was measured, as well as the perceived workload and occurrence of simulator sickness from doing so. The following research questions will be addressed in this chapter:

- **Ch3 RQ1** How is reaction time affected when interacting with a distracting secondary task displayed via AR?
- **Ch3 RQ2** How does the location and presentation of items in an AR display affect performance on a reaction time task?
- **Ch3 RQ3** How does performing a simultaneous AR and cognitive task affect levels of motion sickness?

3.2 Experiment 1

Investigating the Dual-Task Cost of Performing a Secondary Task in AR on Inhibition Control

The first experiment investigated the ability of participants to perform an inhibition control cognitive task while simultaneously performing a secondary task presented via an AR headset. The experimental design was approved through the College of Science and Engineering Research Ethics committee (Application number #300210133). The following section reports the methods, procedure and results of this experiment.

Design

The experiment was designed to compare performance on a Go/No-Go task both with and without performing a secondary AR task at the same time. Furthermore, the presentation of AR content was split into two conditions: Central, where the AR content was overlaid directly on top of the cognitive task and Global, where the AR content was displayed over a larger area requiring participants to move their gaze away from the presentation of the cognitive task. A 2x2 mixed design was used with Group (Central versus Global during dual-task trials) as a between groups factor, and Task type (cognitive only versus AR-cognitive) as a within groups factor. Performance on the Go/No-Go task (correctly responding to Go stimuli and not responding to No-Go stimuli) along with perceived workload and sickness scores were measured as independent variables.

Participants

Forty-two participants *(Mean Age = 27.04 SD = 5.07, 18 Female)* were recruited to take part in the experiment. These were recruited via online forums and around the University of Glasgow Computer Science and Psychology departments. All participants had normal, or corrected to normal eyesight, and were allowed to wear glasses whilst using the headset where required.

Materials

Cognitive Task

The Go/No-Go task as introduced by Nosek and Banaji (2001) was implemented using PsychoPy v2021.2.3 (Peirce et al., 2019) and displayed on an Asus VX279 27-inch monitor approximately 60 cm away from participants. The stimuli consisted of a circle, 10.4cm in diameter, which appeared in the centre of the screen.

The stimuli were on screen for 3000ms at a time, in line with findings from Eriksson and Stanton (2017) for the average length of TOR warnings. Go signals were coloured red (RBG: 255, 0, 0), yellow (RBG: 255, 255, 0) or orange (RGB: 255, 125, 0) in line with Politis et al. (2013), who used similar coloured stimuli to convey a warning to drivers. Participants viewed each stimulus 15 times in a randomised presentation order, and responded by pressing the space bar on a keyboard in front of them. After either a key response from a participant or 3000ms had passed with no response, the stimulus would disappear and the next trial would begin. Between each trial was a period between 5 and 10 seconds, selected randomly, to prevent participants anticipating the start of the trial.

Augmented Reality

The AR task was developed in Unity (version 2020 3.26f1) using the Mixed Reality Toolkit - (MRTK) (version 2.7.2). The Augmented Reality Eye Tracking Toolkit - (ARETT) was used to display and register eye movements (Kapp et al., 2021) in the Microsoft HoloLens 2 AR headset. AR content in the distractor task was designed according to guidelines from Microsoft (2023) and was inspired by the distracting non driving task used by Radlmayr et al. (2018) and the multiple object avoidance task from Mackenzie et al. (2021). Coloured holographic gems would appear in 3D space in front of the participants, who had to manoeuvre a green cursor onto the gems using their gaze. No other feedback was given, but participants were asked to 'pop' all the gems they could see as quickly as possible (see Figure 3.2). The eye-tracking function of the AR task was calibrated using the in-built Microsoft calibration that is required for each new user of the HoloLens 2.

Presentation was divided into two conditions: Global, covering the entire area of the display screen (620mm across, approx. 54.65° visual angle), and Central, where the spawn zone for the gems was constrained to an area approximately 4 times smaller than the whole screen, located on top of where the Go/No-Go stimuli were presented (150mm across, approx 14.25° visual angle). Due to the display size of 52° within the HoloLens 2 (43°×29°), the Global condition required participants to move their head away from where the Go/No-Go stimuli appeared and observe the whole screen to pop all of the gems. Two alignment bars were included in the Unity environment, which were lined up with the edges of the computer monitor by the experimenter prior to beginning the trials to ensure that the AR content was displayed in the correct area. To assess any development of sickness symptoms, the simulator sickness questionnaire Simulator Sickness Questionnaire - (SSQ) (Kennedy et al., 1993) was administered at three separate timepoints: Pre experiment, post control and post experiment. In order to measure perceived workload during the experiment, the NASA Task Load Index - (NASA TLX) (Hart and Staveland, 1988) was administered after both the control and the experimental conditions via the Qualtrics (2024) online questionnaire platform .

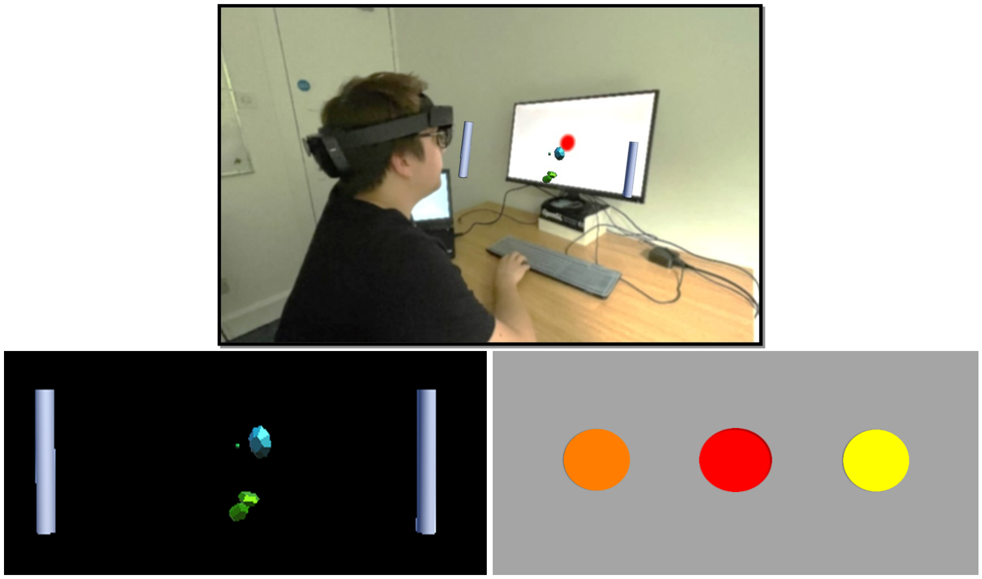


Figure 3.2: Examples of the setup in experiment 1 (top) with participants wearing the HoloLens 2 headset while performing the Go No-Go task. The AR task was to look at and pop the gems that appeared in the headset (bottom left), while responding when the red circle appeared and ignoring the orange and yellow circles (bottom right).

Procedure

Participants were first given an information sheet, demographics questionnaire and consent form to sign. After completion, the procedure of the experiment was explained. Participants first saw a set of six practice trials for the Go/No-Go task to familiarise themselves with the requirements. They were provided feedback on their responses, showing a 'correct' screen if they responded to the Go stimuli/ ignored the No-Go stimuli, or an 'incorrect' screen if they responded to the No-Go stimuli/ignored the Go stimuli. Next, they put on the AR headset and moved onto the experimental trial to perform the cognitive Go/No-Go task only, with nothing appearing in the headset in order to measure their baseline reaction time performance. The only AR objects they saw during this condition were the alignment bars and the eye tracking cursor. During the trials, the Go/No-Go stimuli were shown on the screen for 3000ms and participants had to respond by pressing the space bar as fast and as accurately as possible A buffer period between each trial with a randomly chosen length between 5000ms and 10000ms was used to prevent participants anticipating when the next trial would begin (see Figure 3.3). Next, they were subject to one of the two AR task conditions, either Global or Central. Assignment to one of these conditions was alternated between each participant taking part. The change in performance was measured between the single cognitive tasks and the dual AR-cognitive task over the length of the experiment. Finally, the between group differences between the Central and Global task conditions were also compared. The experiment lasted approximately 30 minutes and participants were provided with a £4 Amazon voucher as compensation for their time.

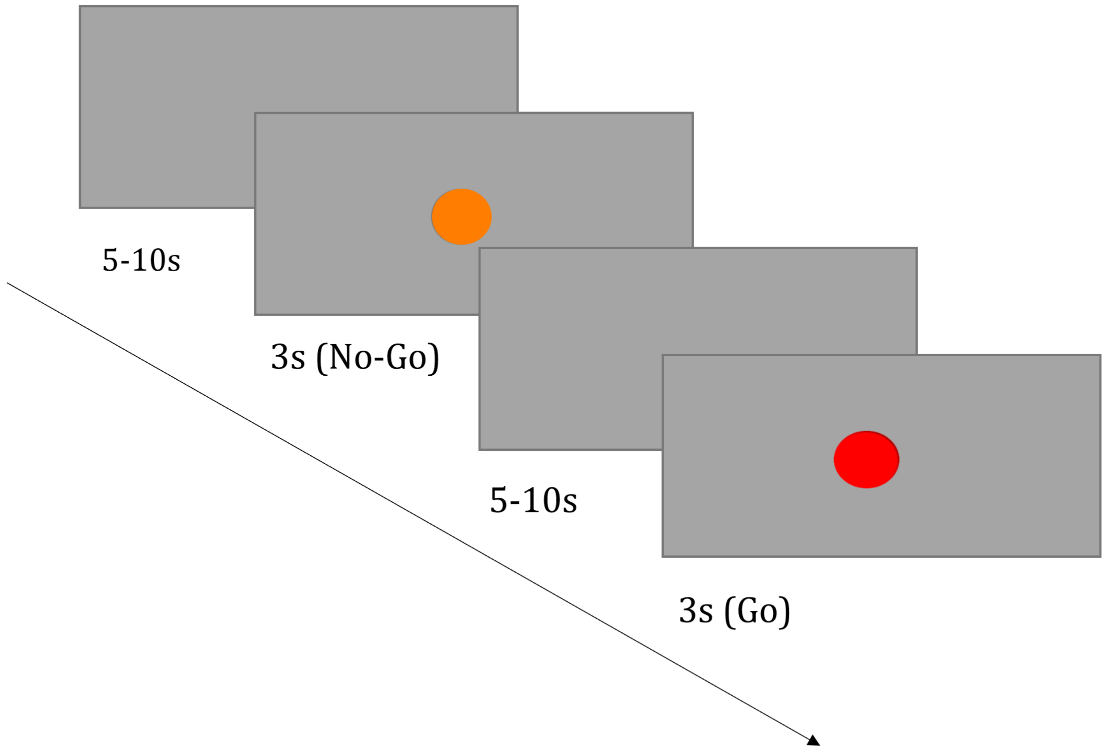


Figure 3.3: A diagram showing the basic trial structure of the Go/No-Go task, with a NoGo (orange) stimuli and a Go (red) stimuli shown for 3 seconds, or until a response was given. Between the presentation of a stimulus was a random period of time between 5 and 10 seconds, to prevent participants anticipating the presentation. Participants first completed only this Go/No-Go task, and then again while simultaneously performing the AR task in the headset.

3.3 Experiment 1 Results

Results were analysed using R Studio 2022.07.01 Build 554 using the lme4 (Bates et al., 2015), lmetest (Kuznetsova et al., 2017) and report (Makowski et al., 2023) R packages. Linear Mixed Effects - (LME) models were fitted to the data and were estimated using Restricted Maximum Likelihood - (REML) and the nloptwrap optimizer. Using LME models was selected due to the repeated measures designs, with multiple responses recorded from the same participants resulting in nested data. To compensate for this effect, the use of mixed effects models were deemed as the most suitable method to account for variation within participants responses within experimental conditions. The analysis techniques set out by Barr et al. (2013) were followed, who suggest a 'keeping it maximal' approach to avoid overcomplicated model fits when using mixed effects models. To achieve this, a step wise model selection approach was used, where models containing sequentially less significant variables are compared and removed, e.g., a model with all possible variables fitted as random intercepts is compared to a model with one of these intercepts removed. The proportion of the variance explained by each model is compared, and models with intercepts which do not significantly explain a greater proportion of this variance are rejected.

This culminates in the retention of the simplest model, with the fewest random intercepts, that explain the greatest proportion of the variance. This maximal model was then also compared to a null model which considers the fixed effects purely as random effects (suggesting they vary randomly and not due to the experimental conditions) again to evaluate whether this model explains a greater proportion of variation than random variance. For a full description, see Barr et al. (2013). Standardised parameters were obtained by fitting the model on a standardised version of the dataset. 95% Confidence Intervals - (CIs) and p-values were computed using a Wald t-distribution approximation.

Reaction Time

The average reaction time of participants on Go stimuli between the baseline cognitive task block was compared to the average reaction time in the dual cognitive-AR task block of the experiment. The data were z-scored and reaction times greater than 3 standard deviations away from the mean were removed as outliers (n = 19). A LME was fitted to predict the interaction effects of Block and Condition on Reaction Time, with the formula:

$$\text{Reaction Time}_{ij} = \beta_0 + \beta_1 \cdot \text{Condition}_{ij} + \beta_2 \cdot \text{Block}_{ij} + \beta_3 \cdot (\text{Interaction})_{ij} + u_{0i} + e_{ij}{}^1$$

The model included participant as random effect and was found to explain significantly greater variance than a null model with Condition and Block fitted as random effects as well as participant *(Null Model AIC = -569.18, BIC = -553.85; Interaction Model AIC = -665.28,BIC = - 634.61; p < .001)*. The total explanatory power was moderate (conditional $R^2 = 0.25$). Within this model, the effect of Task Block was statistically significant *(Est. = 0.11, 95% CI [0.09, 0.14], t(1219) = 7.82, p < .001)* with reaction time increasing in the dual Cognitive-AR task block (see Figure 3.4 & Table 3.1). However, there was no significant effect of AR presentation condition *(p = .773)*.

[1]- Reaction Time$_{ij}$ is the response variable for the i-th observation for the j-th participant. - β_0 is the fixed intercept. - β_1 is the fixed effect coefficient for the Condition variable. - β_2 is the fixed effect coefficient for the Block variable. - β_3 is the interaction effect between Condition and Block. - u_{0i} is the random intercept for the i-th participant, drawn from a normal distribution with mean zero and some participant-specific variance. - e_{ij} represents the residual error term for the i-th observation for the j-th participant.

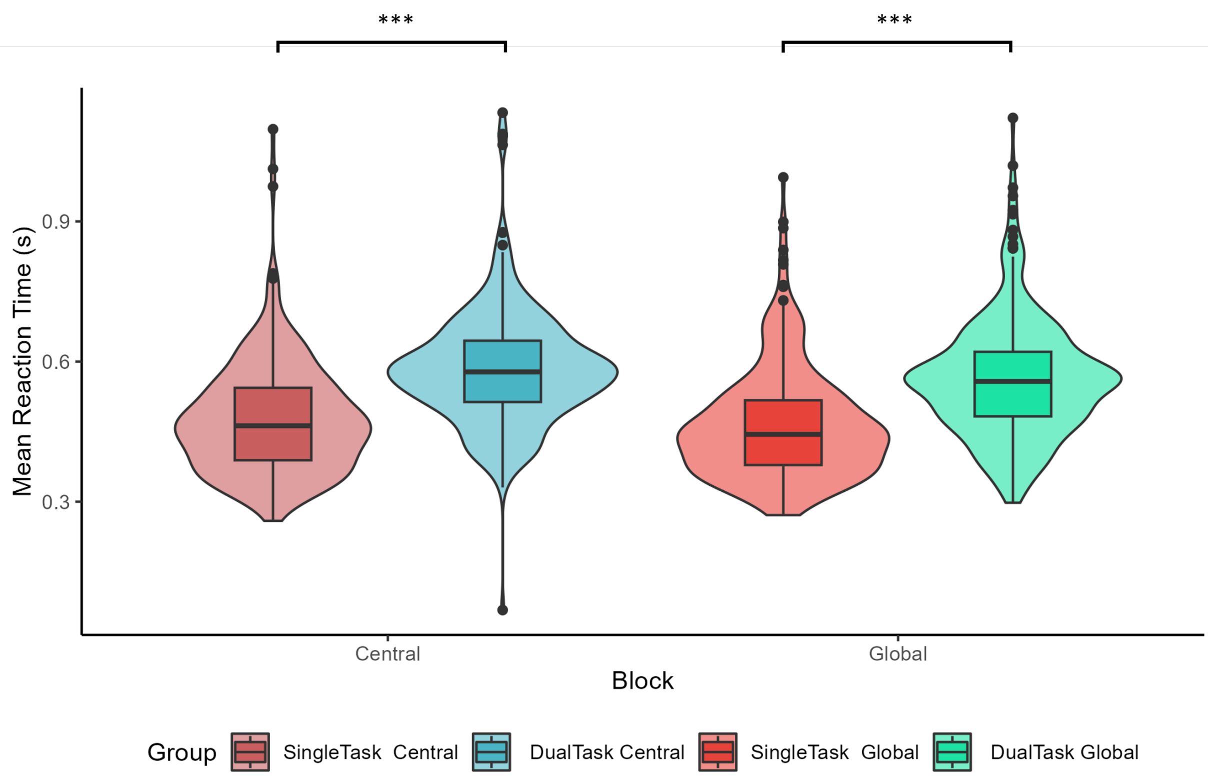


Figure 3.4: Mean reaction time between the Single and Dual Task conditions (bottom left) and the two presentation conditions (bottom right), and the interactions (top). There was a significant difference in reaction time between the task blocks, but not the presentation conditions, nor any interaction effects.

Condition	Block	Mean RT(s)	SD
Central	Cognitive Only	0.49	0.17
Central	Dual Cognitive-AR	0.61	0.22
Global	Cognitive Only	0.48	0.20
Global	Dual Cognitive-AR	0.58	0.19

Table 3.1: Summary statistics for the Go\No-Go reaction times in Experiment 1.

Go/No-Go Performance

To measure proportion of correct responses, the number of Hits, Misses, Correct Rejection and False Alarms were summed and then compared. LME models were fitted to predict the main effect of correct responses with Block, including participant as a random effect. There were significantly more False Alarms *(conditional R^2 = 0.47, beta = -0.29, 95% CI [-0.52, -0.07], t(77) = -2.62, p = .011)* in the dual Cognitive-AR tasks compared to the cognitive only task. However, the models fitted for Hits *(p = .0507)* and Misses *(p = .608)* were not significant (see Figure 3.5).

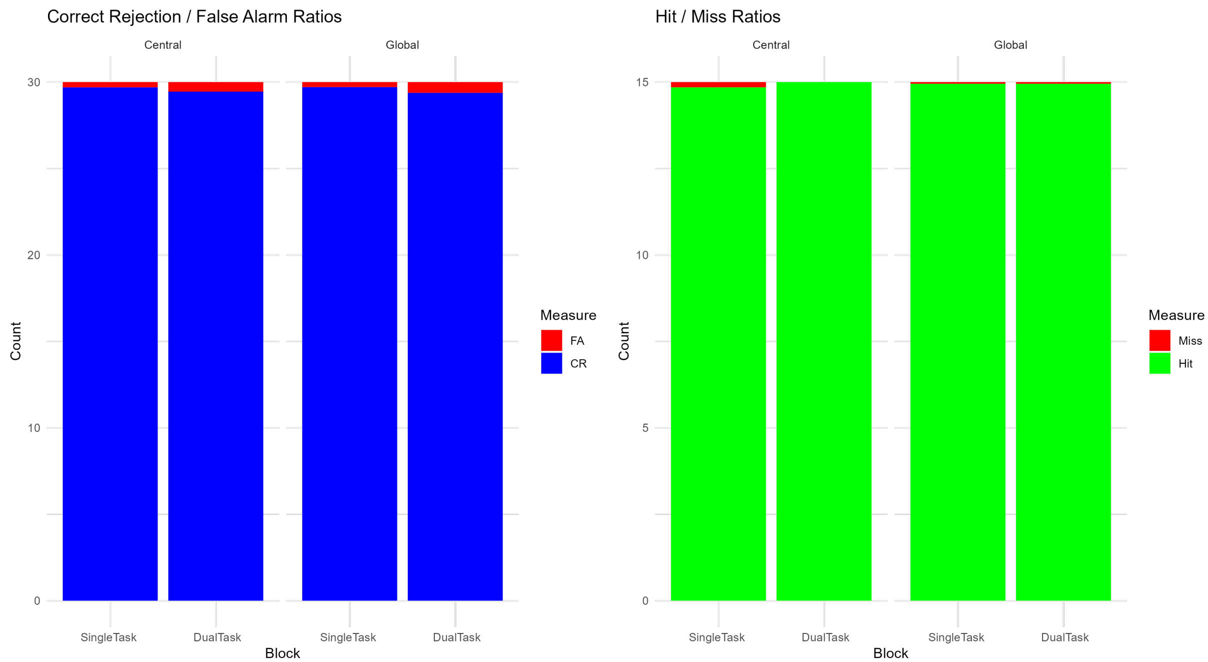


Figure 3.5: Frequency of Hits, Missed, Correct Rejections and False Alarms for the Go/No-Go task.

Simulator Sickness

With the SSQ data, 3 x 2 mixed design ANOVAs were conducted comparing SSQ scores at each of the three timepoints, with timepoint as a within groups effect and participant as a random effect. However, there were no significant differences in sickness scores over the three timepoints *(p = .195)*, nor between the two display size conditions *(p = .125)* on total SSQ score, nor on any of the subscales.

Perceived Workload

A 2 x 2 mixed design ANOVA was conducted comparing the raw total NASA TLX ratings between each condition, with block as a within groups effect and participants as a random effect. The Total TLX rating was statistically significantly different across different conditions $(F(1, 41) = 37.92, p < .001, \eta^2 = 0.18)$, where the Dual task condition was rated as significantly more demanding than the Single task condition. However, there were no significant differences between the AR presentation conditions (see Figure 3.6). Subsequent ANOVAs were run for each of the subscales which also found significant increases in the Dual task condition for all subscales except Frustration (See Table 3.2).

Scale	ANOVA	Sig	η^2
Total TLX Score	$F(1, 41) = 37.92$	**$p < .001$*****	0.18
Mental Demand	$F(1, 41) = 47.73$	**$p < .001$*****	0.21
Physical Demand	$F(1, 41) = 16.27$	**$p < .001$*****	0.14
Temporal Demand	$F(1, 41) = 9.64$	**$p = .003$****	0.1
Overall Performance	$F(1, 41) = 11.25$	**$p < .002$****	0.04
Effort	$F(1, 41) = 25.35$	**$p < .001$*****	0.13
Frustration	$F(1, 41) = 4.01$	$p = .052$	0.01

Table 3.2: Comparisons for the NASA TLX subscales for Experiment 1.

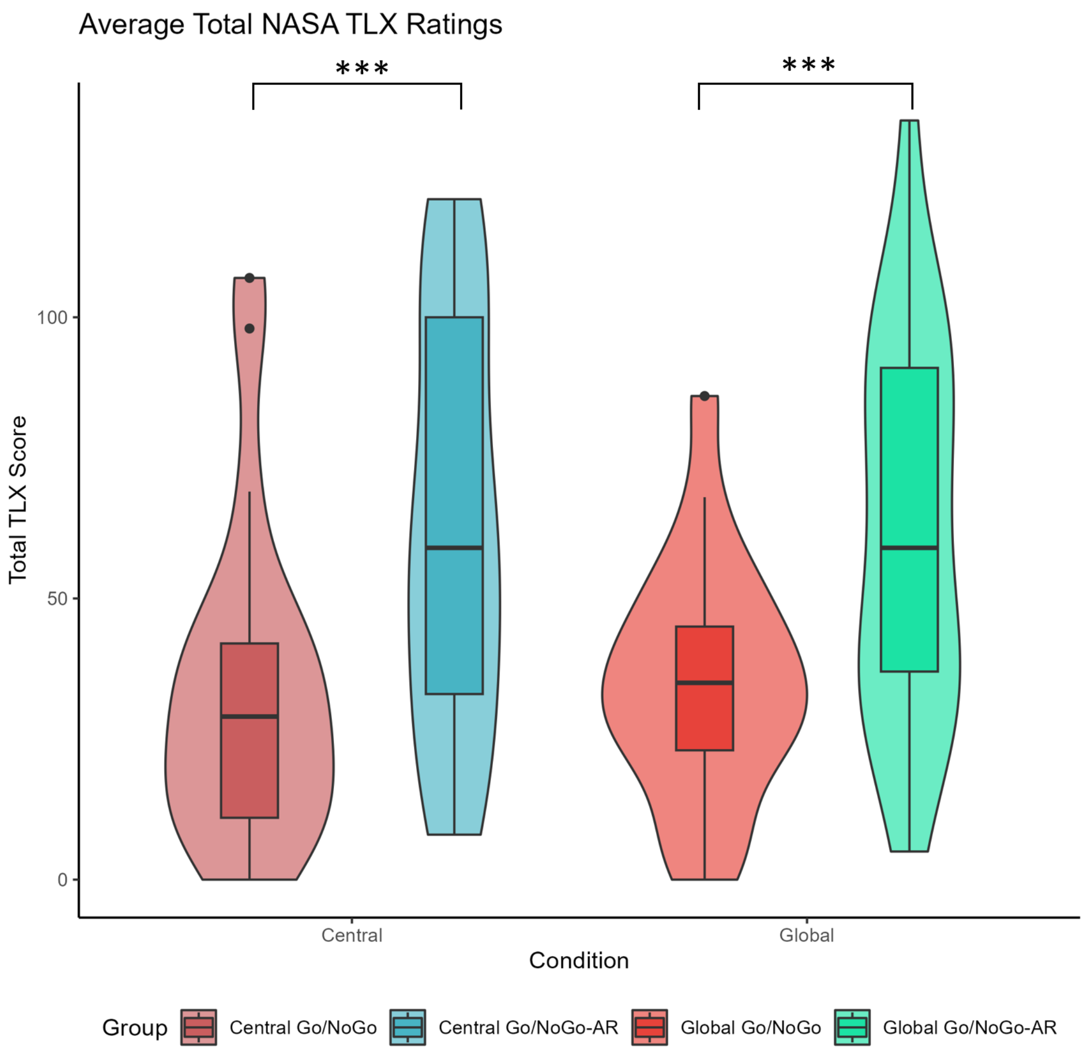


Figure 3.6: Total NASA TLX scores between each condition and each Task Block. There was a significant increase in total TLX scores in the dual cognitive-AR task condition, but no differences between the Global and Central presentation conditions.

3.4 Experiment 1 Discussion

Overall, performance on dual cognitive-AR task led to slower reaction time and increased perceived workload compared to performing a single task. There were also greater occurrences of participants responding to an irrelevant No-Go cue. This is a finding that is concordant with the dual-task literature (Heuer, 1996; Walker et al., 2021; Fereydooni et al., 2019) that splitting attention between tasks leads to a degradation in performance. However, the specific demands of the Go/No-Go task here may also have contributed. Janczyk and Huestegge (2017) demonstrated that in a dual-task paradigm with Go/No-Go as the secondary task, trials with No-Go stimuli actually improved performance on a primary response task. This may have contributed to the results here, where there were twice as many No-Go stimuli as Go stimuli. Participants were required to interact with the AR NDRT more often than the cognitive task, which may have biased attention towards popping the gems when a No-Go stimulus was present. This in turn affected response times to the Go stimuli.

When considering this in an AV context, these results suggest that using an AR display for distracting NDRTs could the reaction time to a TOR, and may promote inappropriate responses to visual alerts. Furthermore, the presence of driver alerts which do not require specific intervention may bias attention towards an NDRT, further impairing the reaction time to alerts that requiring driver intervention. Although the difference in reaction time here is relatively small, a delay in reacting to a TOR means less time afforded for a driver to process the road scene once they regain control. Though the lack of overall sickness is expected from a short exposure to AR, it is appropriate to further investigate how switching between an AR interface and a more complex and dynamic visual task at different depth of field may affect sickness levels.

Whilst this experiment attempted to evaluate a method for measuring performance on dual cognitive-AR tasks, it is not possible to extrapolate the results to a driving setting, due to the simplicity of the tasks and the stimuli. Reacting to the simple Go/No-Go stimuli cannot be extrapolated to the more complex cognitive behaviours required during driving, and the way in which a driver responds to hazards on the road. Therefore, it is necessary to further investigate performance on a cognitive task which is more closely related to the driving task.

3.5 Experiment 2

Investigating the Dual-Task Cost of Performing a Secondary Task in AR on Hazard Perception

To specifically investigate the impact of performing this distracting AR task on driver attention, an experiment was designed to test the demands of differing display size on the performance of simultaneously presented AR and hazard perception tasks. To extend these findings from Experiment 1 into the driving domain, the Hazard Perception test was selected as a probe of driver attention. This test was devised by Horswill and McKenna (2004) to measure the speed and ability of drivers to recognise dangerous road events. A video clip shown from the perspective of the driver is shown, and participants must react as quickly as possible when they notice a developing hazard (defined as an event which "requires the driver to take action, like changing speed or direction" (Driving and Vehicle Standards Agency, 2023). This skill has been reliably linked to the likelihood of being involved in accident (Horswill et al., 2010, 2015) and is used as a measure of a driver's ability to perceive dangers on the road.

An experiment was conducted used comparing performance on either a single HP task, or Dual HP-AR tasks. As in Experiment 1, this AR task was split between two presentation conditions: Central and Global. Participants' performance on the HP task, measured by how much time they took to react to the hazard, was compared as a baseline to their performance whilst also conducting the AR task at the same time. The difference in HP performance between the two AR presentation conditions and the baseline was also compared. In this instance, the Central condition was intended to emulate how experienced drivers' attention tends to rest on the focus of expansion on the road, the point where motion originates from in the visual field (Crundall and Underwood, 1998; Chapman and Underwood, 1998). Whereas, the Global condition encouraged a broader search pattern encouraging participants to look around the whole visual scene. Again, the SSQ (Kennedy et al., 1993) was administered at three separate time points: Pre experiment, post baseline and post experiment. The NASA TLX (Hart and Staveland, 1988) was also administered after both the Single and Dual task conditions. The experimental design was approved through the College of Science and Engineering Research Ethics committee (Application number #300210313).

Clips	Hazard	Length	ClipOnset	ClipOffset	HazardWindow	Hazard Location
Clip1	Van brakes hard for turning vehicle ahead	0:36:11	22.18	25.00	2.82	Centre
Clip2	Car pulls put from right with cyclist on left	0:45:10	34.43	40.00	5.57	Right
Clip3	Pedestrians step into road from the right	0:49:03	22.45	27.00	4.55	Right
Clip4	Car door opens from left	0:25:11	13.94	19.00	5.06	Left
Clip5	Car turning left stops in road	0:29:51	12.42	16.00	3.58	Left
Clip6	Car turning left has to stop in road	0:41:45	29.72	33.00	3.28	Left
Clip7	Pedestrian steps into road from the right	0:57:22	34.91	40.00	5.09	Right
Clip8	Cyclists in the road	0:57:45	41.93	47.00	5.07	Centre-right
Clip9	Van encroaches on your lane	0:33:53	16.82	20.00	3.18	Centre
Clip10	Pedestrians step into the road from the left	0:42:26	31.75	36.00	4.25	Left
Clip11	Car starts reversing into the road	0:41:46	22.00	26.00	4.00	Centre-right
Clip12	Cyclist encroaches on your lane	0:35:38	26.60	29.00	2.40	Centre-left
Clip13	Van pulls off without indicating	0:33:35	20.8	24.00	3.20	Centre-left
Clip14	Pedestrian runs into the road from the right	0:19:30	9.59	12.00	2.41	Right
Clip15	Van pulls out from the left	0:38:07	16.82	20.00	3.18	Centre-left
Clip16	Pedestrian steps out from the left	0:23:08	12.78	15.00	2.22	Centre-left
Clip17	Van pulls out from the left	0:24:25	11.1	15.00	3.90	Centre-left
Clip18	Pedestrian steps out from the left	0:45:01	27.91	33.00	5.09	Left
Clip19	Car stopped in the middle of the road	0:31:12	24.20	27.00	2.80	Centre
Clip20	Pedestrian steps out from the left	0:19:30	8.00	13.00	5.00	Left/Centre
Clip21	Car door opens from left	0:41:08	22.66	28.00	5.34	Left
Clip22	Pedestrians in middle of road	0:32:37	20.72	25.00	4.28	Centre
Clip23	Multiple cars pulling across your path	0:29:37	15.82	17.00	1.18	Centre
Clip24	Taxi pulls into your lane	0:46:37	35.48	39.00	3.52	Centre-right

Table 3.3: Summary details for all of the Hazard Perception clips, with their lengths, onset time and the location of the hazard in each clip.

Creating the Hazard Perception Clips

To create the HP clips, a GoPro Hero 360 Max camera was attached to the windscreen of a Citroen C3 car to capture the road from a driver's perspective. Footage from around the Greater Glasgow area was collected between 10am and 4pm from March until June.

Filming routes consisted of a mixture of urban, suburban, and rural roads. In total, 10 hours of footage was filmed, which was then reviewed and edited by a trained traffic psychologist to identify hazards, following the criteria set out by Goodge et al. (2021). The definition for a hazard was taken from the UK Government's Hazard Perception test as *"something that would cause you to take action, like changing speed or direction"* (Driving and Vehicle Standards Agency, 2023). A scoring window was defined for each clip, from the first point the hazard appeared in the video until after it was no longer on screen (see Table 3.3 for the full list of clips).

Participants

Twenty-eight participants *(Mean age = 36.96 SD = 9.92, 11 Female)* were recruited via online forums and around the University of Glasgow Computing Science and Psychology departments. They were required to have held a driving license for at least two years to be eligible *(Mean driving experience = 15.43. SD = 10.21)*. Nineteen participants held a UK driving license, with other participants reporting licences from Germany (3), USA (2), Brazil (1), China (1), Italy (1), Russia (1) and Turkey (1).

Figure 3.7: A picture of the driving simulator setup. Participants viewed the Hazard Perception clips on the centre screen whilst wearing the headset, and responding by pressing one of the buttons on the steering wheel.

All participants had normal or corrected to normal vision and were allowed to wear glasses whilst using the headset, where required. Fourteen drivers reported having taken part in a Hazard Perception test previously. Four participants reported having experience with AR, eight reported having some experience with AR and sixteen reported having no experience with AR at all.

Procedure

Participants were presented with the HP clips whilst sat in a driving simulator setup, with the HP task presented on the front screen and the AR task presented in the Microsoft HoloLens 2 (Figure 3.7). In keeping with the official presentation of the Hazard Perception test, the side monitors were turned off to show a single screen and reduce potential distraction. First, participants viewed 10 randomly selected hazard clips to measure their baseline reaction time performance. They were instructed to press a button on the steering wheel as soon as they spotted a hazard in the video clip. As above, this was defined to participant as an event which would *"require you as the driver to take action, like changing speed or direction"*. At the end of this 10-clip block, they provided SSQ and NASA TLX ratings. For the second block, this procedure was then repeated with the remaining 10 hazard clips that had not been seen, whilst also performing the AR NDRT. Participants were told to fixate on as many gems as possible to make them disappear, while also focusing on the HP clips. Following this block, further NASA TLX and SSQ ratings were taken.

3.6 Experiment 2 Results

Results were analysed using the same procedure as Experiment 1 (see section 3.3).

Hazard Perception

The average reaction time between the baseline HP task block was compared to the average reaction time in the dual HP-AR task block of the experiment. A LME model was fitted to predict the main effects of Reaction time with Block. The model included participant and hazard clip as random effects and was found to explain significantly greater variance than a null model with Condition and Block fitted as random effects as well as participant *(Null Model AIC = 1444.8 , BIC - 1457.2; Interaction Model AIC = 1028, BIC = 1057.1; p < .001)*. The model's total explanatory power was substantial (conditional $R^2 = 0.31$), and had the formula:

$$\text{Reaction Time}_{ij} = \beta_0 + \beta_1 \cdot \text{Condition}_{ij} + \beta_2 \cdot \text{Block}_{ij} + \beta_3 \cdot (\text{Interaction})_{ij} + u_{0i} + u_{0j} + e_{ij}$$ [2]

Within this model, the effect of Block was statistically significant *(Est. = 0.26, 95% CI [0.08, 0.43], t(461) = 2.91, p = .004)*, with reaction time significantly increasing overall in the dual HP-AR conditions (see Figure 3.8). This suggests there was a significant increase in reaction time on the HP task when participants performed the AR task simultaneously, compared to just the hazard perception task (See Table 3.4). However, there were no significant differences between the AR presentation conditions *(p = .078)*, nor any interaction effects *(p = .914)*. A LME model was also fitted comparing the number of responses between each Block and each Condition. However, no significant differences were found for any of the comparisons.

Condition	Block	Mean RT(s)	SD
Central	Baseline HP	1.68	1.43
Central	HP-AR	1.92	1.56
Global	Baseline HP	1.39	1.29
Global	HP-AR	1.60	1.15

Table 3.4: Summary statistics for the hazard perception reaction times in Experiment 2.

[2]- Reaction Time$_{ij}$ is the response variable for the i-th observation for the j-th participant. - β_0 is the fixed intercept. - β_1 is the fixed effect coefficient for the Block variable. - β_2 is the fixed effect coefficient for the Condition variable. - β_3 is the interaction effect between Block and Condition. - u_{0i} is the random intercept for the i-th participant, drawn from a normal distribution with mean zero and some participant-specific variance. - u_{0j} is the random intercept for the j-th Clip Number, drawn from a normal distribution with mean zero and some Clip Number-specific variance. - e_{ij} represents the residual error term for the i-th observation for the j-th participant.

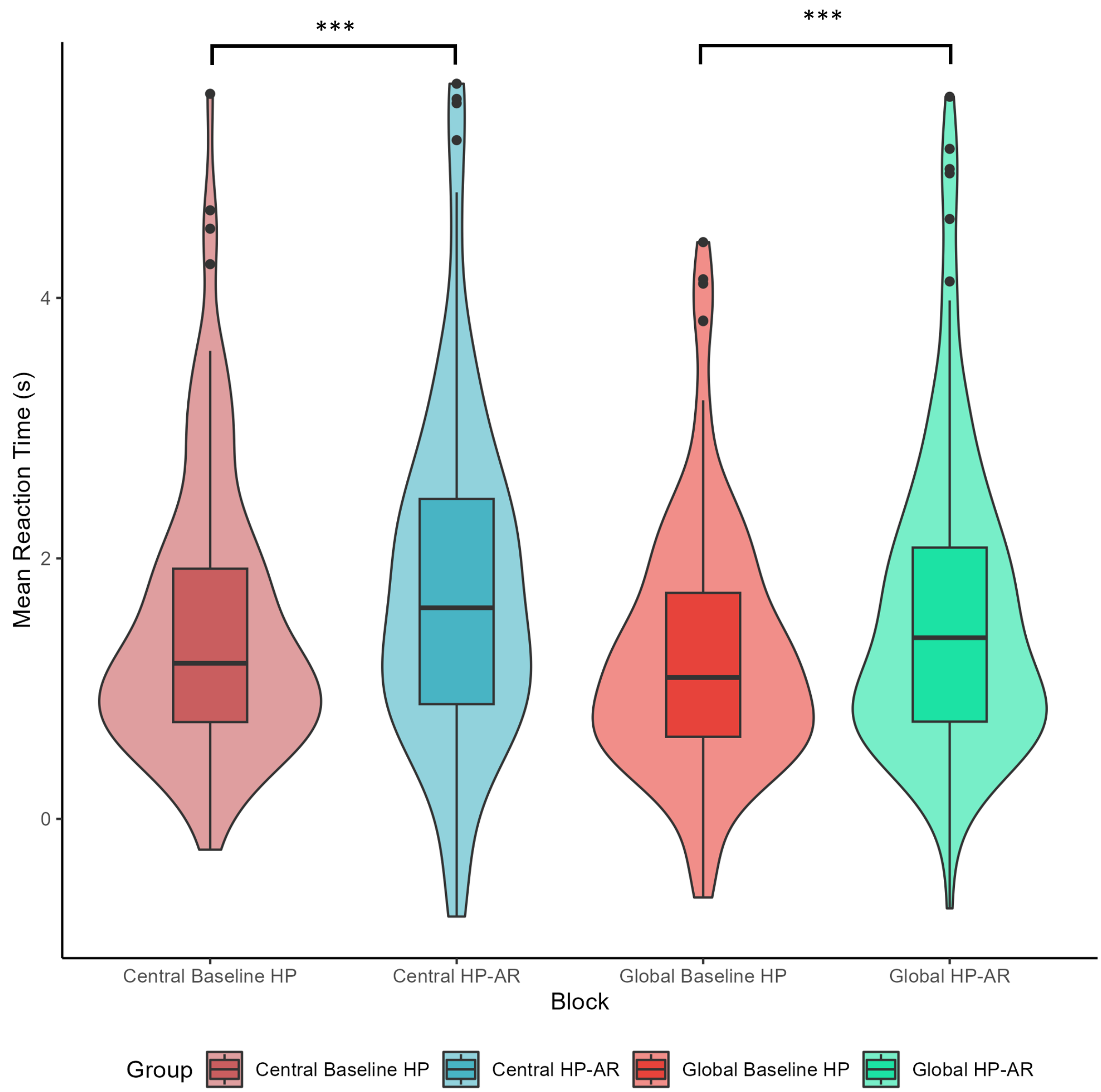


Figure 3.8: Change in average reaction time between the baseline HP only and HP-AR dual task conditions Overall (triangle) as well as for each AR presentation condition (Central as circles and Global as Squares). There was a significant increase in reaction time on the HP task when participants performed the AR task simultaneously, compared to just the hazard perception task. However, there was no significant difference between the AR presentation conditions.

Simulator Sickness

A 3 x 2 mixed design ANOVA was conducted comparing Total SSQ scores at each timepoint of the experiment with timepoint as a within groups effect and participants as a random effect, but there were no significant differences were found in Total SSQ scores. These scores were then converted to show change over the different timepoints of the experiment, where the score for each timepoint was subtracted from the previous one with the baseline kept at 0. An increase in reported sickness from 0 to 1 would give a positive value, and a decrease from 1 to 0 would give a negative value, showing change in SSQ scores over time.

After converting the scores between the Baseline, HP-only and HP-AR blocks, a 3 x 2 mixed design ANOVA was conducted to predict SSQ scores at each Timepoint was fitted, again with timepoint as a within groups effect and participants as a random effect. This showed there was a significant increase in Total SSQ scores $(F(1.18, 30.64) = 7.52, p = .007, \eta^2 = 0.13)$. *Post hoc* comparisons with a Bonferroni correction revealed significant increases between the HP-AR condition and both the Baseline ($p = .031$) and HP only $(p = .019)$ conditions.

A 3 x 2 mixed design ANOVA comparing changes in the Oculomotor subscale of the SSQ found a significant change in ratings $(F(1.43, 37.26) = 9.83, p = .001, \eta^2 = 0.16)$, with *post hoc* comparisons with a Bonferroni correction showing significant increases for the HP-AR condition compared to both the HP only $(p = .01)$ and Baseline $(p = .006)$ conditions. Finally, another 3 x 2 mixed design ANOVA comparing changes in the Disorientation subscale also found significant differences $(F(1.42, 36.92) = 7.88, p = .004, \eta^2 = 0.14)$, with *post hoc* comparisons with a Bonferroni correction showing significant increases for the HP-AR condition compared to both the HP only $(p = .015)$ and Baseline $(p = .02)$ conditions (see Figure 3.9).

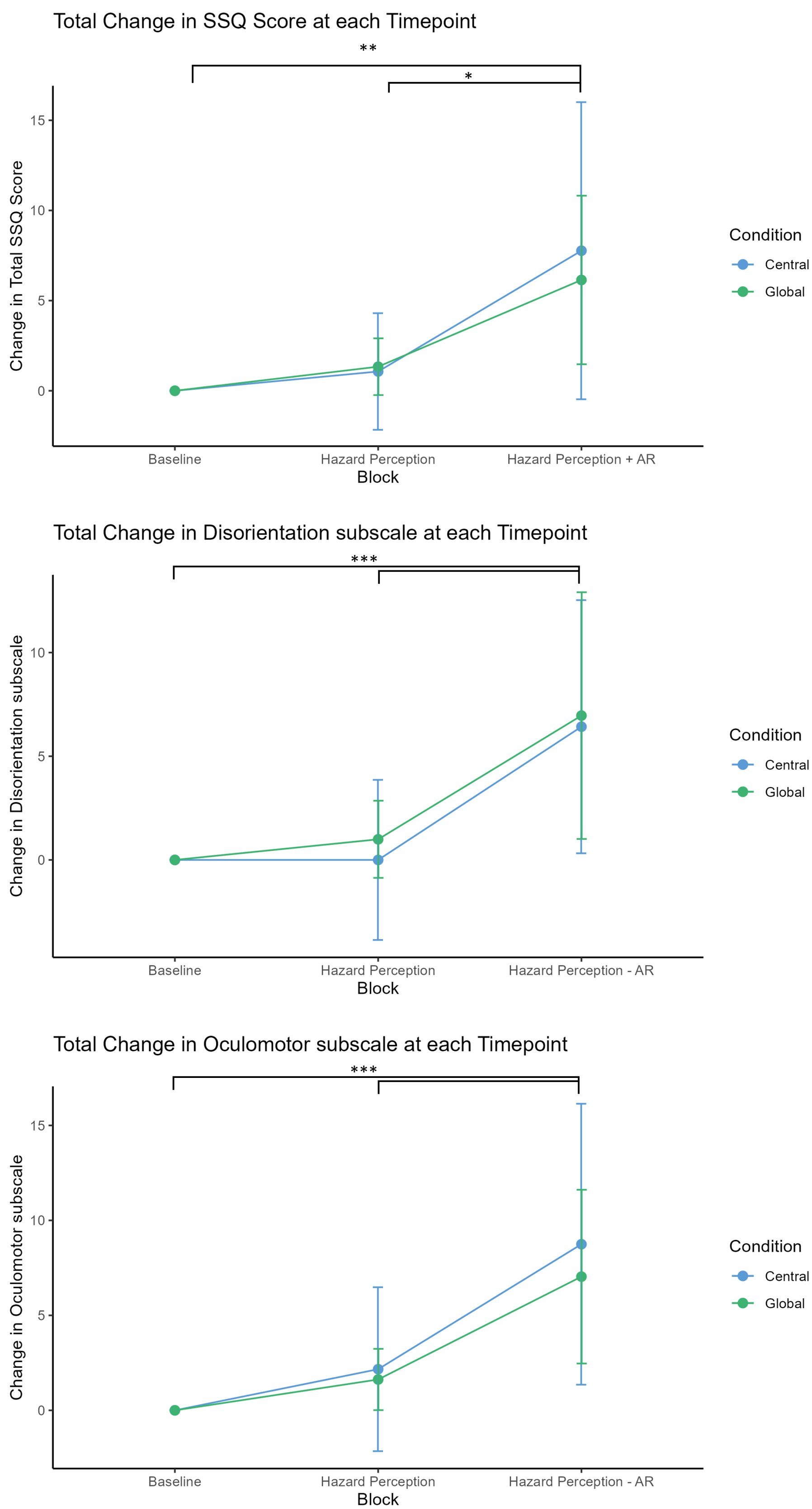


Figure 3.9: Change in SSQ scores over each timepoint of the experiment. There were significant increases in the HP-AR block of the experiment for the Total SSQ score (top left), as well as for the Disorientation (bottom left) and Oculomotor (bottom right) subscales.

Perceived Workload

A 2 x 2 mixed design ANOVA was conducted comparing the raw total NASA TLX ratings between each condition, with block as a within groups factor and participant as a random effect. The Total TLX rating was statistically significantly different across different conditions *(F(1, 23) = 30.51, p < .001,* η^2 *= 0.24)*, where the Dual task condition was rated as significantly more demanding than the Single task condition. However, there were no significant differences between the AR presentation conditions (see Figure 3.10). Subsequent ANOVAs were run for each of the subscales which also found significant increases in the Dual task condition for all subscales (see Table 3.5).

Scale	ANOVA	Sig	η^2
Total TLX Score	F(1, 23) = 30.51	**p < .001*****	0.24
Mental Demand	F(1, 23) = 43.06	**p < .001*****	0.29
Physical Demand	F(1, 23) = 16.6	**p = .001*****	0.17
Temporal Demand	F(1, 23) = 21.98	**p = .003****	0.17
Overall Performance	F(1, 23) = 17.92	**p = .002****	0.25
Effort	F(1, 23) = 10.77	**p < .001*****	0.11
Frustration	F(1, 23) = 12.18	**p = .02***	0.11

Table 3.5: Comparisons for the NASA TLX subscales for Experiment 2.

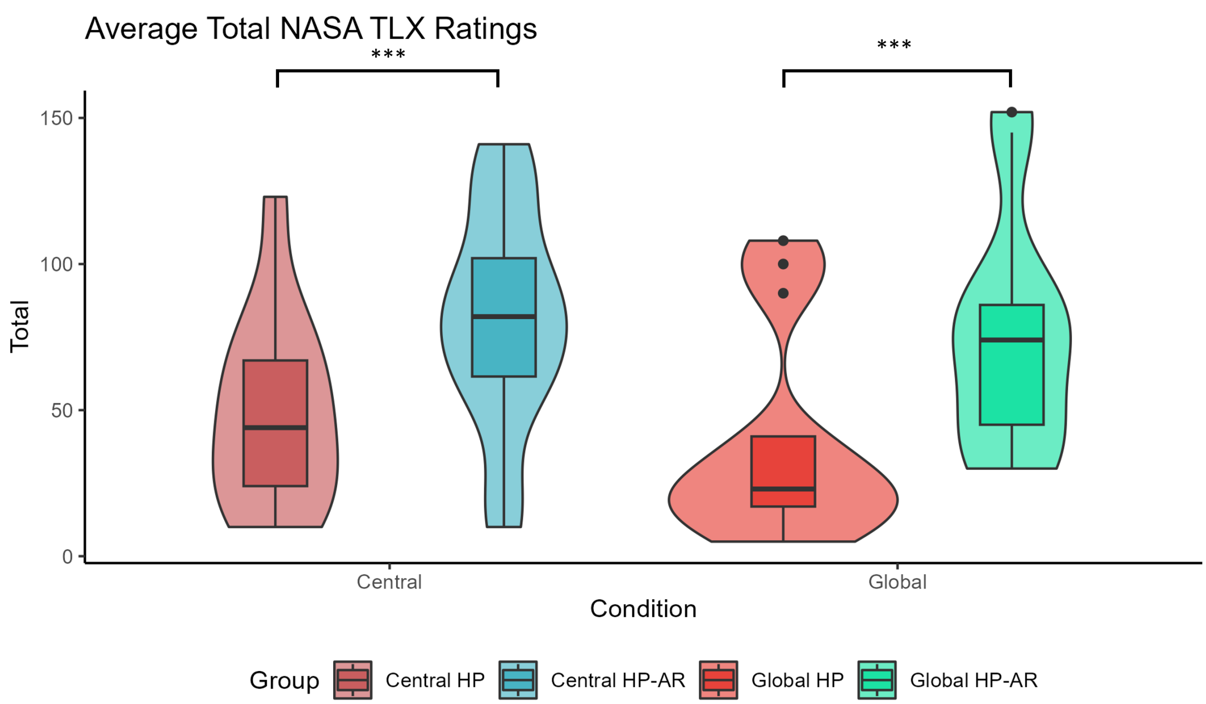


Figure 3.10: Change in NASA TLX scores between the HP and HP-AR conditions.

3.7 Chapter 3 Discussion

The results from the experiments presented in this chapter indicate that using an AR interface impairs performance on a simultaneous reaction time task. Reaction time to both the target stimuli in Experiment 1 and to hazardous events in Experiment 2 was significantly slower in the dual-task conditions with a distracting AR task when compared to single task performance. These results are in line with previous literature, which shows how dual task paradigms results in poorer performance (Heuer, 1996; Jackson et al., 2023). Overlaying the secondary task in AR onto the primary task does not mitigate this effect, indicating that participants were impaired by having to switch task from the AR task to the Cognitive tasks. This pattern of results was also found in a study also measuring this simultaneous performance on an office AR task and a Tetris game, which showed that switching between tasks was not an instantaneous process (Fereydooni et al., 2019). A similar delay as a result of task-switching was seen here. While a 0.5 second delay is not a substantial amount of time, when considered in a driving context it can be a significant amount of time. This 0.5 second delay in a driving context translates to a vehicle moving at 60 mph travelling an extra 7 metres before the driver reacts to the hazard, half of the thinking distance needed should the driver need to brake (RAC, 2023).

This is a finding echoing the literature where performing multiple tasks impacts driver performance (Walker et al., 2021; Ward and Helton, 2022), hence why non-driving tasks are illegal in manually controlled vehicles. Taken into an AV context however, allowing drivers to engage with an NDRT will split their attention between the supervision task and the NDRT, resulting in a delayed reaction to any information about the road (Ou et al., 2021b) or a TOR (Meiran, 2000). Here, these experiments showed that performing an NDRT led to a small yet still significant increase in reaction time. This suggests that interacting with an NDRT via AR interface may also interfere with the underlying vigilance task of maintaining awareness in an AV. Whilst previous studies have investigated the feasibility of AR display during the supervision of AVs (Riegler et al., 2019b, 2022), there has been little research which specifically measures the impact of using an AR interface for an NDRT. However, the AR task used here are not wholly representative of the types of tasks that drivers might engage with when supervising an AV. A survey by Wilson et al. (2022) found that leisure activities were reported as the preferred NDRT to engage with, followed by sleep and socialising. The AR task used here was designed to disrupt the fixation patterns that experienced drivers exhibit and is similar to AR tasks used in other research investigating NDRTs (Muguro et al., 2021). However, it can not be said to be fully representative of the full range of NDRTs that drivers may engage with, nor the demands these may put on driver's attention.

Furthermore, it is difficult to make general conclusions about driver awareness from the results presented here. The hazard perception task used to measure driver behaviour is not wholly representative of the driving task. Whilst it has merits in testing reaction times to hazardous events, as well as propensity for involvement in accidents (Deery, 1999; Horswill and McKenna, 2004; Horswill et al., 2015), it does not involve the complex coordination between reaction time, motor control and appraisal needed when in control of a vehicle. Additionally, it is subject to a criterion bias of what an individual's perception of a hazard is, how dangerous it is and when it starts to occur (Pradhan and Crundall, 2016; Crundall, 2016). Participants were instructed to carry out the hazard perception task and so attention on the clips formed the main part of the task. Whereas, a driver supervising an AV is less likely to be as engaged in the supervision task (Gruyer et al., 2017). Prior research has established the benefits of drawing the driver's attention to dangers in the road, and how this provides a benefit to attention (Rusch et al., 2013; Stefanucci et al., 2022; Wu et al., 2023b). Though the results presented here contribute to those which have established the impact of simultaneously performing an NDRT on cognitive performance, the potential for using AR as a more dynamic and informative display (e.g., Schömig et al. or Schroeter et al.) requires further exploration.

In particular, methods which can tap into the driver's situational awareness of the road scene, such as the Situational Awareness Global Assessment Tool - (SAGAT) test (Endsley et al., 1998) can provide a better way of assessing driver awareness (Crundall, 2016; Gugliotta et al., 2017; Ventsislavova and Crundall, 2018) and how this might be impacted by an NDRT. This test involves freezing a HP video just before testing the knowledge of the participants on their understanding of the scenario. Using the SAGAT method, (Kim and Gabbard, 2022) found that driver SA was improved through the use of an AR HUD but also proved distracting in other circumstances, depending on what was displayed via the HUD. Furthermore, testing these driver behaviours in a more valid environment such as a driving simulator or on the road is a more effective method to draw conclusions about how drivers might interact with HUDs in AVs. (Gabbard et al., 2019) found that a 3D AR interface encouraged longer glances onto the road and was rated as being more demanding. Conversely, Jing et al. (2022) found that AR HUDs were able to reduce driver distraction, but this was only while watching videos of driving scenarios and without engaging with an NDRT. Though the Hazard Perception method presented here is used as a measure of a driver's ability to notice hazards in the road, it is not wholly representative of the way drivers process a road scene. Methods such as the SAGAT which measures a driver's awareness of a road scene at a higher cognitive level would provide further insight into the impact on driver awareness an NDRT displayed in AR would have.

Cybersickness in AR

Despite extensive attention in the Virtual Reality domain (McGill et al., 2017; Rangelova and Andre, 2018; Pöhlmann et al., 2022a), measuring cybersickness in AR has not received the same research interest. The increase of SSQ scores after the dual cognitive-AR tasks is concordant with previous studies which measured sickness in AR (Hughes et al., 2020; Pettijohn et al., 2020). In particular, the increase in the disorientation subscale but not the nausea subscale is consistent with the distinction between cybersickness and motion sickness (Stanney et al., 1997) and is similar to the results found by (Kaufeld et al., 2022). One potential reason is the conflicting visual cues may increase discomfort, similar to what has been found in VR (Dam and Jeon, 2021). Here, although the hazard perception clips lasted no longer than a minute and the exposure to AR was brief, there was still a significant increase in SSQ scores over time in Experiment 2, as well as in the disorientation and oculomotor discomfort subscales. Combined with the reported increase in perceived workload when performing the AR task, there needs to be further exploration into the type of AR task that can be presented in these dual-task scenarios which do not cause detrimental effects on the user. Additionally, the AR NDRT was world locked and remained stable during the task. This is in contrast with the HP clips showing near-constant motion cues or *vection* via the movement of the vehicle in the video. Vection, or the perception of motion while stationary (Duh et al., 2004), has been linked to increased sickness in VR environments (Bonato et al., 2008; Li et al., 2021) and has also been shown to lead to increased sickness levels when viewing HP clips in a VR environment (Goodge et al., 2021). Similarly, the frequent switching between different depths of the AR task and real-world screen may have led to the increased eye strain (Gabbard et al., 2018), but previous research also suggests that the relationship between depth and sickness is unclear (Lisle et al., 2019). Future research should include measures to evaluate how much AR-induced sickness is affected by the content displayed or individual differences, e.g., familiarity with computer games (Häkkinen et al., 2006; Bigoin et al., 2010).

3.8 Conclusions

The experiments in this chapter measured the effects of interacting with an NDRT presented as an AR display on a primary cognitive task. Reaction time on both an inhibition control and a hazard perception task was impaired when participants also performed a secondary AR task, compared to baseline single-task performance. To address **RQ1** of this thesis (section 1.3), these results indicate that performing a secondary NDRT leads to poorer reaction time. This is consistent with previous literature investigating dual task performance. Whilst there has been a wealth of research investigating how AR displays could be used to present driving-related information, the results here suggest that presenting an NDRT to drivers in AR impairs their ability to respond to the road. Furthermore, regarding **RQ2** of the thesis, the different size of the AR displays showing the NDRT did not have different impacts on performance, with similar detriment evident in both the displays. The results from this chapter demonstrate the impact of presenting NDRTs in AR on driver cognition and reactions times, which needs to be considered when designing interfaces for displaying NDRTs. However, questions remain over the impact of AR displays on the ways in which drivers process and understand the road scene beyond a simple reaction time measure, which the Hazard Perception test does not capture. The next chapter outlines an alternative method for measuring driver awareness, along with the creation and validation of such a method used for measuring *situational awareness* empirically, which is used in the subsequent chapters that address question of using cues to aid driver awareness posed by **RQ3**.

Chapter 4

Creating a Hazard Prediction test as an Empirical measure of Situational Awareness

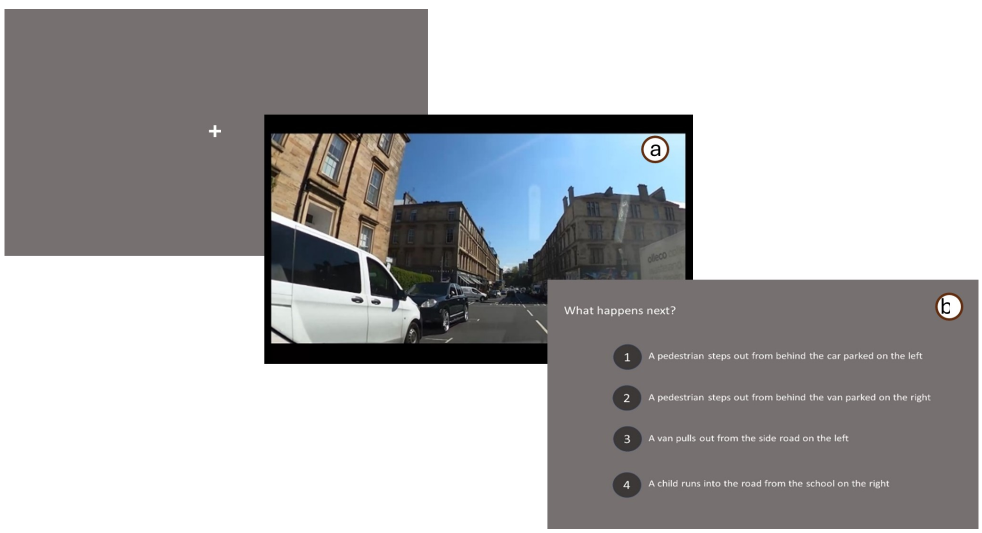


Figure 4.1: The stages of viewing a single clip in the Hazard Prediction test presented in this chapter.

This chapter describes the creation and validation of a Hazard Prediction test, a variant of the Situational Awareness Global Assessment tool, as a means of empirically measuring driver awareness. This test involves participants viewing a video clip of the road filmed from the point of view of the driver. Each clip contains a single hazardous road event, which requires action from the driver. Just before the hazard occurs, the clip is occluded and participants are asked to predict what they think happens next from a list of multiple-choice options. An experimental and statistical validation of this test are presented in this chapter, justifying its use to probe situational awareness in Experiments 3, 4, 5, and 6. It is described in detail here to be referred to in later chapters.

4.1 General Introduction

Measuring Driver Perception

The ability to perceive dangers on the road is an important part of determining one's skill as a driver. To make safe and appropriate decisions, a driver must be able to cognitively process complex road scenes while also operating the vehicle. Measuring this ability, however, has proved to be a non-trivial task for researchers. As mentioned in the previous chapter, the preeminent method, as pioneered by Horswill and McKenna (2004), uses a video-based test where a video filmed from the perspective of a driver is presented and participants must press a button when they notice any dangerous road event. This *Hazard Perception* test saw success in subsequent years as a means of distinguishing between more and less safe drivers (Horswill et al., 2015). In particular, Hazard Perception skill has been shown to predict involvement in collisions (Horswill et al., 2010, 2015) and to effectively discriminate between older and younger drivers (Borowsky et al., 2010). The Hazard Perception test allows insight into how younger drivers cognitively process the road (Moran et al., 2020), providing an explanation as to why they are over-represented in collision data (Department for Transport, UK Government, 2021). Consequently, the UK, Dutch, and Australian Governments employ a version of the Hazard Perception test as a requirement for gaining a driving license. Here, the time when the hazard is visible is split into windows, with participants gaining more points on the test if they react earlier towards the start of the window, rather than towards the end. Their score is then compared to a threshold in order to pass part of a standardised theory test.

However, when trying to gain an empirical measure and understand the cognitive state of drivers, psychologists have identified a number of limitations with the Hazard Perception test. Different drivers will define a hazardous road event based on their own subjective thresholds and criteria. The UK Government defines a Hazard as *"something that would cause you to take action, like changing speed or direction"* (Driving and Vehicle Standards Agency, 2023) for their Hazard Perception test, but this definition can change depending on the driver (Deery, 1999). Typically, an experienced driver will react to a developing hazard much earlier than a novice, as evidenced by fewer overall incidences of harsh braking and acceleration (Cao et al., 2014). Using the hazard window scoring, this means experienced drivers do not score points on the test, as they can react 'too early' before the valid hazard window despite having predicted its occurrence, and thus are penalised. Not only does individual perception of hazards differ, different driving cultures have different definitions of what a 'hazard' is. A complex and busy road scene in the UK which is challenging to the most experienced of driver could be seen as routine in a country like Malaysia (Lim et al., 2013).

This affects both reaction times to each hazard (Lim et al., 2013; Ventsislavova et al., 2019; Di Stasi et al., 2020) and cultural perception of what a hazard is (Ventsislavova et al., 2019; Lim et al., 2014) requiring each Hazard Perception test to be specific to the country it is administered in. Furthermore, the Hazard Perception test is ostensibly a measure of reaction time. Yet there are other key attributes of driver cognition when they are processing a road scene.

Situational Awareness - (SA), the ability to maintain up-to-date attention of their current environment, has been identified as an important part of a driver's ability to navigate the road safely (Endsley, 1995a; Chaparro et al., 1999; Crundall, 2016). In her seminal work, Endsley (1995a) defined the model of SA to involve three distinct processes: 1) Perception of elements in the environment 2) Comprehension of their current state and importance and 3) Projection of what is likely to happen to these elements next. Applied to driving, this theory has been used to try and explain the cognitive processes that drivers go through when monitoring the road. In order to navigate a road scene safely, a driver must be able to process what is in their environment (other vehicles, pedestrians, road signs and markings), understand what the current situation is (is a vehicle coming towards or away from you, traffic conditions, a pedestrian crossing, lane restrictions) and then predict what will happen next (will the vehicle ahead suddenly brake, are the pedestrians about to cross, will the road layout ahead change). By processing all of these factors and then predicting other road user's behaviour, a driver is able to make decisions to navigate through a road scene safely. This has been suggested as an important factor for hazard avoidance (Pradhan and Crundall, 2016) and facilitating smoother driving with less harsh braking and steering (Li et al., 2017; Yang et al., 2021b). However, the Hazard Perception test, as it is currently utilised, is not able to tap into these higher level cognitive processes. It is primarily used to discriminate between more or less safe drivers based on their reaction time within a set window, meaning that other measures are needed to probe awareness.

One such example has been proposed based on the Situation Awareness Global Assessment Tool (Endsley et al., 1998). Here, a participant's current knowledge of a road scene is measured by suddenly occluding their view of a simulation or video. Participants demonstrate their SA by describing things like what they have seen, where the location of the hazard is, and what they think happens next in the video clip. If they are able to correctly answer these three questions, this suggests they have comprehensive SA of the scene. This provides an effective way of measuring the higher cognitive processes of drivers with an objective measure: whether a participant's prediction was accurate or not. This is in contrast to the Situational Awareness Rating Technique - (SART), another method developed by Endsley et al. (1998) where operators provide a self-assessment of their SA.

Though this is a useful subjective measure of a person's perceived SA and is easy to employ, the SART suffers from the bias from an operator's own perception of their performance which is affected by workload and perceived success (Endsley et al., 1998). The objective the scoring mechanism of the SAGAT avoids this.

A driving specific version of the SAGAT, known as the Hazard Prediction or 'What Happens Next' (WHN) test, was first proposed by Jackson et al. (2009). They compared an occlusion-based test with video clips that cut to black immediately prior to the hazard onset, with a freeze frame version where the final frame of the clip remained on screen. Both experienced and novice drivers were given a free response to answer what they thought happened next, and how confident they were in their answers. Jackson et al. (2009) found that experienced drivers performed significantly better on the occlusion test than novices, whose performance only improved when given sufficient time to process the freeze frame version. From this, Jackson et al. (2009) posited that this question-based version of the hazard perception test is better able to probe the higher SA levels that experienced drivers use to make appropriate decisions on the road. Developing on this idea, Crundall (2016) showed how this occlusion-based test was able to effectively discriminate experienced drivers from their novice counterparts based on the ways in which they process the road scene. Experienced drivers were shown to effectively use precursor clues to the hazard, such as an indicator flashing or a change in road position, to inform their predictions. Crundall (2016) found experienced drivers consistently performed better than novices, regardless of clip length, the point at which the clip was occluded, or the type of hazard. Ventsislavova and Crundall (2018) refined the paradigm by using free responses to the Hazard Prediction test to create a multiple-choice version of the test and provide a simple scoring mechanism for prediction accuracy: correct or incorrect. Participants choose from a list of 4 options given as potential scenarios to a previous free-response Hazard Prediction test, selecting the option they believe matches to what is about to happen next. Ventsislavova et al. (2019) found there was no significant difference in performance between the multiple-choice and free response versions.

This multiple-choice version of the Hazard Prediction test has been shown to be applicable to a wide range of driving scenarios. It has been applied to fire and rescue drivers in both fire cars (Kroll et al., 2020) and fire appliances (Crundall and Kroll, 2018), as well as with judging the width of a gap for larger fire appliance vehicles (Kroll and Crundall, 2019). The Hazard Prediction test has also seen benefits as an online video format combining hazard perception, prediction and theory questions (Crundall et al., 2021), as well as reducing cybersickness when implemented into a virtual reality domain (Goodge et al., 2021). However, one key advantage of the Hazard Prediction test is that it has been shown to be culturally agnostic (Ventsislavova et al., 2019; Ventsislavova Petrova, 2019).

Cultural expectations and criteria for what a *'hazard'* consists of differ between driving cultures, which affects their responses times to a hazard perception test and their gaze behaviour (Di Stasi et al., 2020). As described previously, a busy road junction which might be quite demanding of a typical UK driver, would be commonplace for a driver from a country such as Malaysia with a more demanding everyday driving environment (Ventsislavova et al., 2019). This in turn affects different driving cultures responses to a Hazard Perception test (Di Stasi et al., 2020). Ventsislavova et al. (2019) specifically compared performance on both Hazard Perception and Hazard Prediction tests of drivers from China, UK, and Spain. Significant differences in response times between the different cultures were identified, dependent on which country the hazard clip was filmed in. However, when using the Hazard Prediction version of the same clips, there were no significant differences between scores for these cultures. This was attributed to the objective nature of the scoring mechanism of this test which does not rely on a subjective interpretation of whether a road event is identified as a hazard or not. This is key for developing a universal test which probes a driver's awareness of the road environment in an empirical fashion which does not rely on subjective judgements of how hazardous a situation is.

While the benefits of using the Hazard Prediction test to probe driver awareness are evident, creating the test is not a simple task. Many hours of footage needs to be collected, reviewed, edited and compiled into an experimental format before the test can be used. Furthermore, deciding on occlusion points for the hazard clips is also a non-trivial task. Crundall (2016) found that manipulating the occlusion point of Hazard Prediction clips led to different results between experienced and novice drivers, with earlier cut times disadvantaging more junior drivers. There needs to be enough information evident about the hazard for participants to recognise what happens next, but not too much as to make it obvious to deduce, invalidating the test. As such, each clip requires reviewing so that each occlusion point is suitable for each hazard. Additionally, in order to evaluate the effectiveness of a Hazard Prediction test, it must be able to sufficiently discriminate between Novice and Experienced drivers (Crundall, 2016). Novice drivers do not possess the experience on the road that allows them to form the accurate predictions of experienced drivers. A valid hazard prediction test needs to include sufficient anticipatory cues in the video clips before the hazard onset so that drivers can use their prior driving experiences to aid their awareness of the situation. The rest of this chapter describes the creation and validation of a version of the Hazard Prediction test made up of a set of 40 hazard clips. An online validation study comparing Experienced and Novice drivers on their Hazard Prediction ability was conducted to determine whether this version of the test sufficiently discriminated between driver groups. Following this, k-fold cross validation was conducted to predict a driver's group using their scores to evaluate the effectiveness of each clip.

4.2 Validation Experiment

Validating the Hazard Prediction Clips Ability to Discriminate between Experienced and Novice Drivers

Stimuli

A GoPro Hero 360 Max camera was attached to the interior windscreen of a Citroen C3 car to capture the road from the driver's perspective (see Figure 4.2). This was placed to capture the forward perspective of the driver on a medium lens setting. Footage from in and around the Greater Glasgow area was collected at various times throughout the day between March and June. Routes included a mixture of urban, suburban, rural and motorway driving to collect a comprehensive representation of different hazards from different road environments. These were conducted between 9.30 am and 5 pm on days without any rain. Twelve hours of footage was collected in total, which was reviewed by a trained traffic psychologist to identify any potential hazards. The definition for a hazard was taken from the UK Government's Hazard Perception test as *"something that would cause you to take action, like changing speed or direction"* (Driving and Vehicle Standards Agency, 2023). To identify a hazard, the same principles as Goodge et al. (2021) were applied:

- The danger present required the driver to change behaviour (e.g., speed, lane position) to prevent a possible collision
- The hazard had a precursor clue (e.g., a car approaching the give way line in a side road ahead)
- The hazard had an easily defined onset (e.g., a car crossing the give way line of a side road and pulling out)
- The hazard should be caused by another road user and not the driver of the film car
- The scene prior to the hazard onset contained other plausible causes of a hazard that allowed distractor options to be generated

The hazard occlusion was chosen at the point just before action from the camera car driver was required, which was deduced from viewing their driving behaviour in the video. These hazards were not staged beforehand but naturally encountered on the road whilst filming. (see Appendix Table A.1 for full list of hazards).

The hazard clips were edited using Adobe Premiere Pro 2023 into clips between 10 seconds and 1 minute in length, and letterboxing was applied to obscure parts of the camera view that were not important, i.e., the dashboard which appeared across the bottom of the video, and the equivalent amount of sky. Each hazard was assigned a description to be given as the correct multiple-choice option to choose from. The other 3 false scenarios were created by looking at the last few frames in each clip and creating plausible but definitively incorrect answers, based on other vehicles or road features visible in the clip. For instance, if the hazard in a clip was 'pedestrians stepping out from behind a parked car from the left', an example false answer could be regarding a) the position of the hazard: *'pedestrians step out from a parked car from the right'*, b) the subject of the hazard: *'a cyclist pulls in front of you from the left'* or c) another vehicle/road feature altogether, depending on the scene: *'the white van pulls into the road from the right'* (see Appendix Table A.1 for the full list of Hazard Clips and their descriptions).

Participants

A power analysis was conducted using the *pwr* package (Champely et al., 2016) in R Studio, which determined that for a moderate effect size of 0.8, 24 participants were required in each group. The experimental sample consisted of 30 Experienced drivers *(Mean Age = 41.5, Mean Driving Experience = 20.2 years, Female = 14)* and 30 Novice drivers *(Mean Age = 30.2, Mean Driving Experience = 0.97 years, Female = 12).* Drivers were considered to be Novices if they held a provisional driving license and had not yet passed their driving test. Some Novices reported more than 2 years driving experience, but had not gained a full driving license and so were still classified as Novices. All participants reported the UK as their country of residence with the majority reporting British nationality, not including 2 Experienced drivers (1 Nigeria, 1 Singapore) and 8 Novice drivers (2 Ireland, 1 Nigeria, 1 Portugal, 1 Latvia, 1 Poland, 1 China).

Procedure

After consenting to take part in the study, participants completed a questionnaire recording their demographics, driving experience and the ratings of the Driver Behaviour Questionnaire - (DBQ) (Reason et al., 1990), presented on the Qualtrics online questionnaire platform (Qualtrics, 2024). This scale asks drivers to rate their propensity to engage with certain driving behaviours, categorised into Errors, Lapses or Violations. Following this, they were automatically directed to the experiment, built using PsychoPy (Peirce et al., 2019) (v2021.2.3) and hosted using the online experiment platform Pavlovia.

Figure 4.2: a) The drivers view from the camera vehicle with the GoPro mounted on the windscreen to the left to emulate the drivers view. b) An example of one of the Hazard Prediction clips that this camera view provided.

Participants first underwent a practice trial to familiarise themselves with the format of the clips and means to give a response. After this, participants saw the 40 hazard clips in a randomised order, broken down into 4 blocks of 10 clips with optional breaks in between. After watching each clip, the multiple-choice list of 4 potential scenarios were presented, of which one was correct. Participants selected the scenario they thought was likely to happen next using the numbers 1-4 on their keyboard (Figure 4.3). After responding to all 40 clips, participants were redirected back to Prolific to receive their £7 compensation.

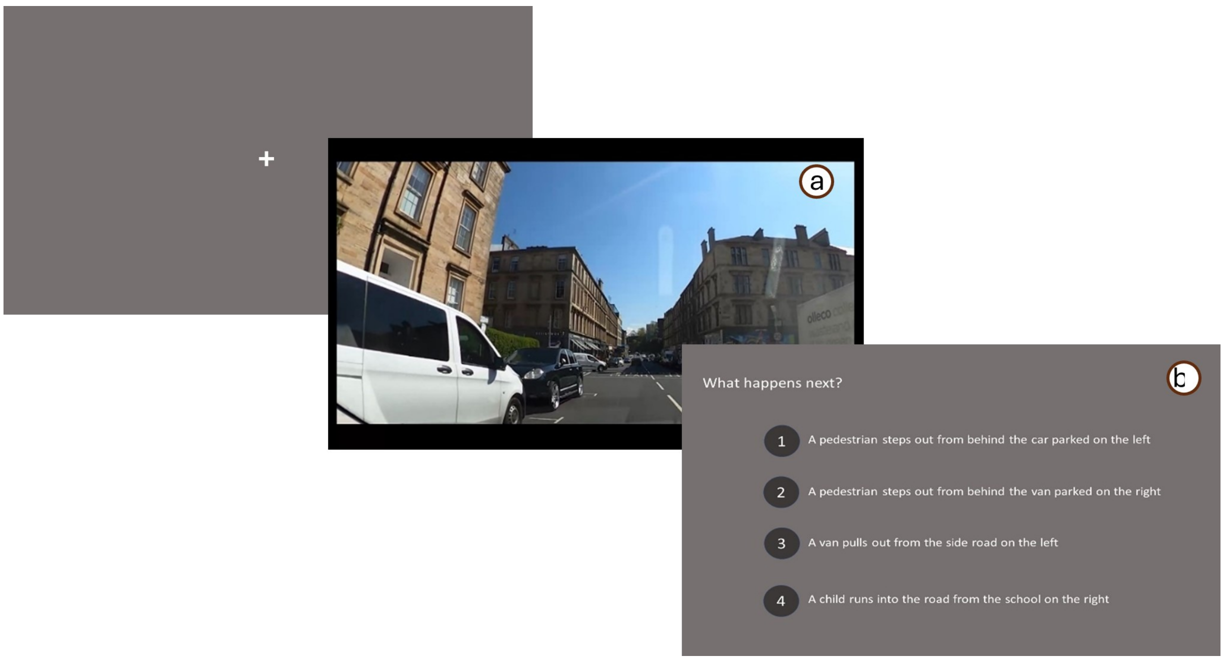


Figure 4.3: a) A screenshot of one of the Hazard Prediction clips, with the Hazard “A van pulls out of the road on the left” visible. b) After watching these clips, the video would occlude and participants would be presented with the 4 multiple-choice options.

4.3 Validation Experiment Results

The results were analysed using R Studio 2022.07.01 Build 554 using the lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017) and report (Makowski et al., 2023) R packages. The responses of each participant were modelled as binomial correct/incorrect distributions in order to calculate the probability of a participant providing a correct response for each of the 40 Hazard Clips. A Generalised Linear Mixed Effects - (GLME) model was fitted to the data for Hazard Prediction score estimated using Maximum Likelihood - (ML) and the bobyqa optimizer. Confidence Intervals - (CI) of 95% and p-values were computed using a Wald t-distribution approximation.

Hazard Prediction Score

Performance on the Hazard Prediction test was compared between Experienced and Novice drivers. A GLME model was fitted with moderate explanatory power (conditional $R^2 = 0.13$) to predict Correct response by Group, with participant as a random effect with the formula:

$$\text{Correct Response}_{ij} = \beta_0 + \beta_1 \cdot \text{Group}_{ij} + u_{0i} + e_{ij}{}^1$$

Within this model, there was a significant difference in the number of correct answers between Experienced and Novice drivers *(Est. = -0.82, 95% CI [-1.17, -0.46],* $p < .001$*)* (see Figure 4.4). Models including the main effect of the country participants obtained their license, years of driving experience and whether participants had experience driving in Glasgow were also fitted, but these were found not to significantly provide any greater explanation of the data and so the simpler model was retained.

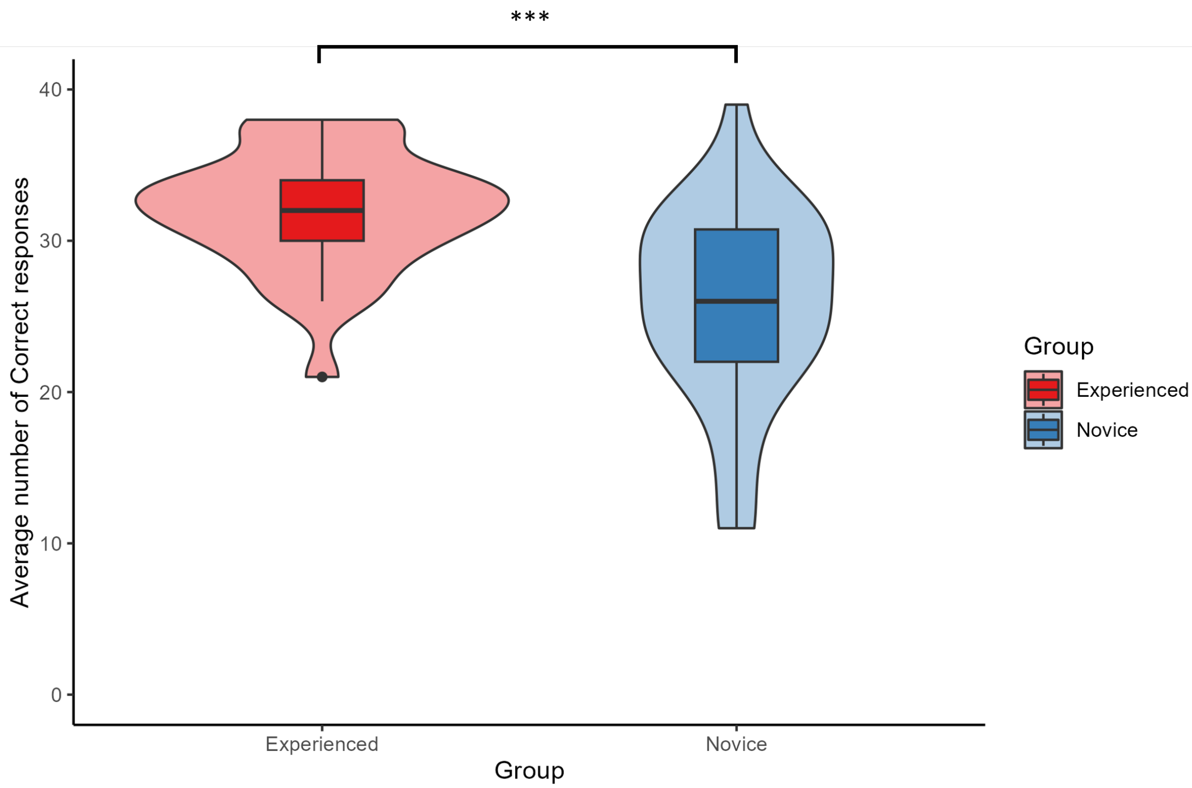
Driver Behaviour Questionnaire

Linear models were fitted to measure differences between the two groups on the 3 subscales of the DBQ, Lapses, Errors and Violations. There were no significant differences between experience group for the Errors or Lapses subscales on the DBQ. There was however a significant difference for Violations, with Experienced drivers reporting significantly higher number of violations than Novices *(conditional* $R^2 = 0.19$*, Est. = -0.51, 95% CI [-0.79, -0.23],* $t(58) = -3.66$*,* $p < .001$*)*. The Experienced drivers were then median split into High and Low groups based on their scores on the Violations subscales. However, models fitted comparing High and Low violations group found that there was no significant difference between these groups on their hazard prediction scores.

Clip by Clip Analysis

Although there were significant differences between scores for each driving experience group, this only provides a general insight into the ability of the test to discriminate between Experienced and Novice drivers. To further validate this test, it was deemed necessary to evaluate performance on each individual hazard clip.

[1]- Correct Response$_{ij}$ is the response variable for the i-th observation for the j-th participant. - β_0 is the fixed intercept. - β_1 is the fixed effect coefficient for the Group variable. - Group$_{ij}$ is the value of the Group variable for the i-th observation for the j-th participant. - u_{0i} is the random intercept for the i-th participant (Participant), drawn from a normal distribution with mean zero and some participant-specific variance. - e_{ij} represents the residual error term for the i-th observation for the j-th participant. The fixed effects are denoted by β coefficients, and the random effects are represented by u terms.



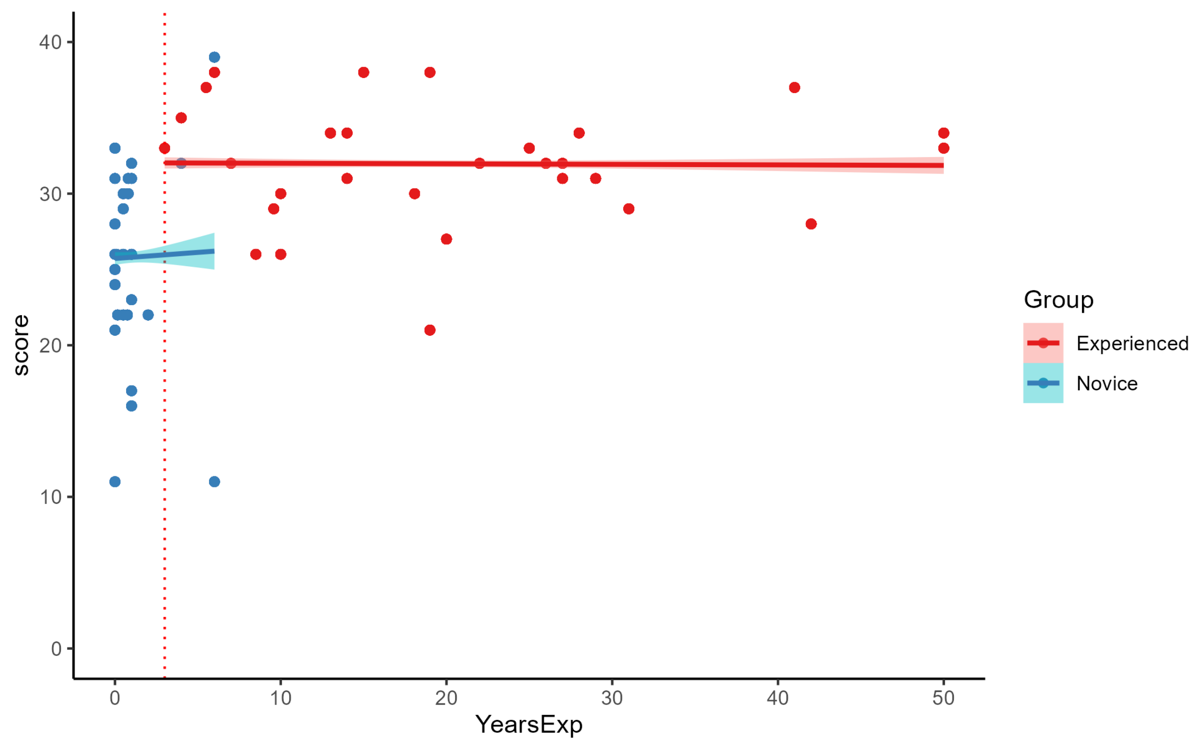


Figure 4.4: (Top) Average number of correct responses for the Experienced and Novice groups. Experienced drivers performed significantly better on the Hazard Prediction test compared to Novice drivers. (Bottom) Reported years experience driving with the score for each driving group. The red dotted line indicates the 3 years driving experience cutoff requirement for the Experienced group. Some Novice drivers reported more than 3 years driving experience, but it was confirmed that they had not qualified for their drivers license yet.

To do this, a ridge regression model was created using the glmnet (Friedman et al., 2010) package to predict a driver's experience group based on their response to each clip (Jackson, 2024). Correct scores were again modelled as a binary response vector (1 for correct, 0 for incorrect). Ridge regression was used due to the high theorised collinearity between responses to each clip i.e., if a driver from the Experienced group responds correctly to one clip, they should also be able to respond correctly to other clips. Using this method reveals the discriminatory power of each clip that makes up this Hazard Prediction test, modelling each clip's ability to predict driver experience group as a coefficient.

First, the most appropriate tuning parameter λ which minimises the mean standard error was identified as 2.2097 (see Figure 4.5). K-fold cross validation was then conducted with k = 10 folds, meaning that for 60 total participants, there were 5 groups of 10 participants used as training data and 1 group of 10 participants as testing data. This was repeated through 100 iterations to account for randomness and produce an average reliability score. With each iteration of the k fold, the mean prediction accuracy of the regression model was calculated, resulting in a distribution of 100 mean accuracies. These were then used to produce an overall mean prediction accuracy of the test of 0.65 *(SD = 0.02, min = 0.6, max = 0.7)* across the 100 iterations, indicating that the model was able to predict driver experience based on their responses to the hazard prediction clips at a greater than chance level (see Figure 4.7).

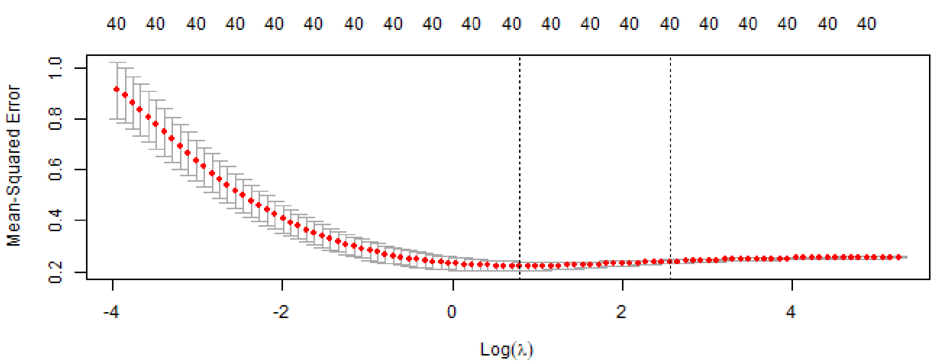


Figure 4.5: A graph showing the best λ selection for the ridge regression model at 2.209667.

From this predictive model, the contribution of each clip was extracted as a predictive coefficient. Each coefficient represents the change in log odds from 0 or no predictive value. The greater the difference from 0, the greater the effect of the individual clip in the model predictions (see Figure 4.6). A higher coefficient value indicates the discriminative power of a clip, meaning that there was a greater difference in performance between Experienced and Novice drivers. The full list of coefficients for each clip, along with the average score for Experienced and Novice drivers is presented in Table B.1 in the Appendix.

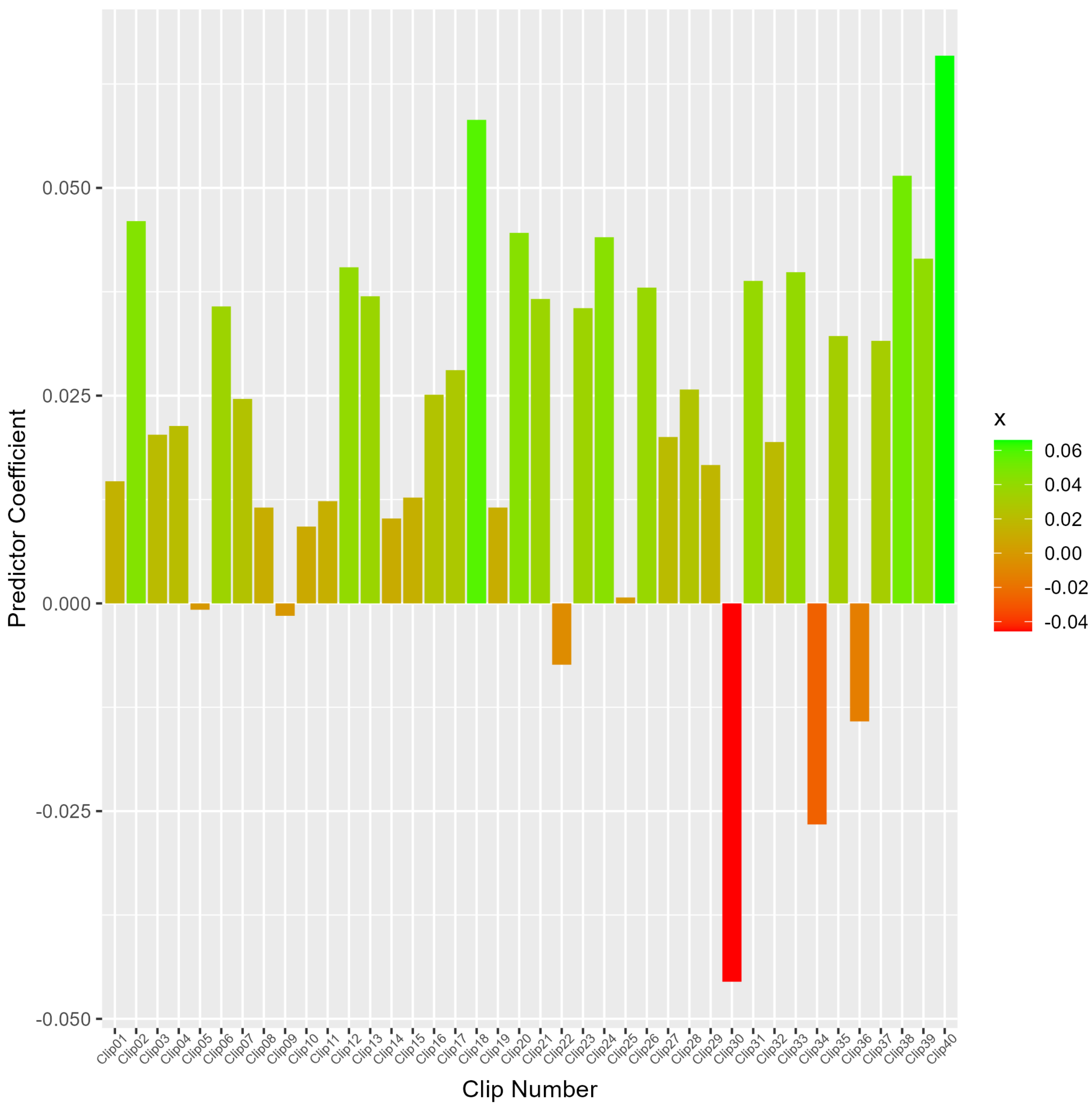


Figure 4.6: Prediction coefficients for each clip on the ridge regression model. The majority of clips showed a positive prediction coefficient, which meant they were useful for discriminating between Experienced and Novice drivers.

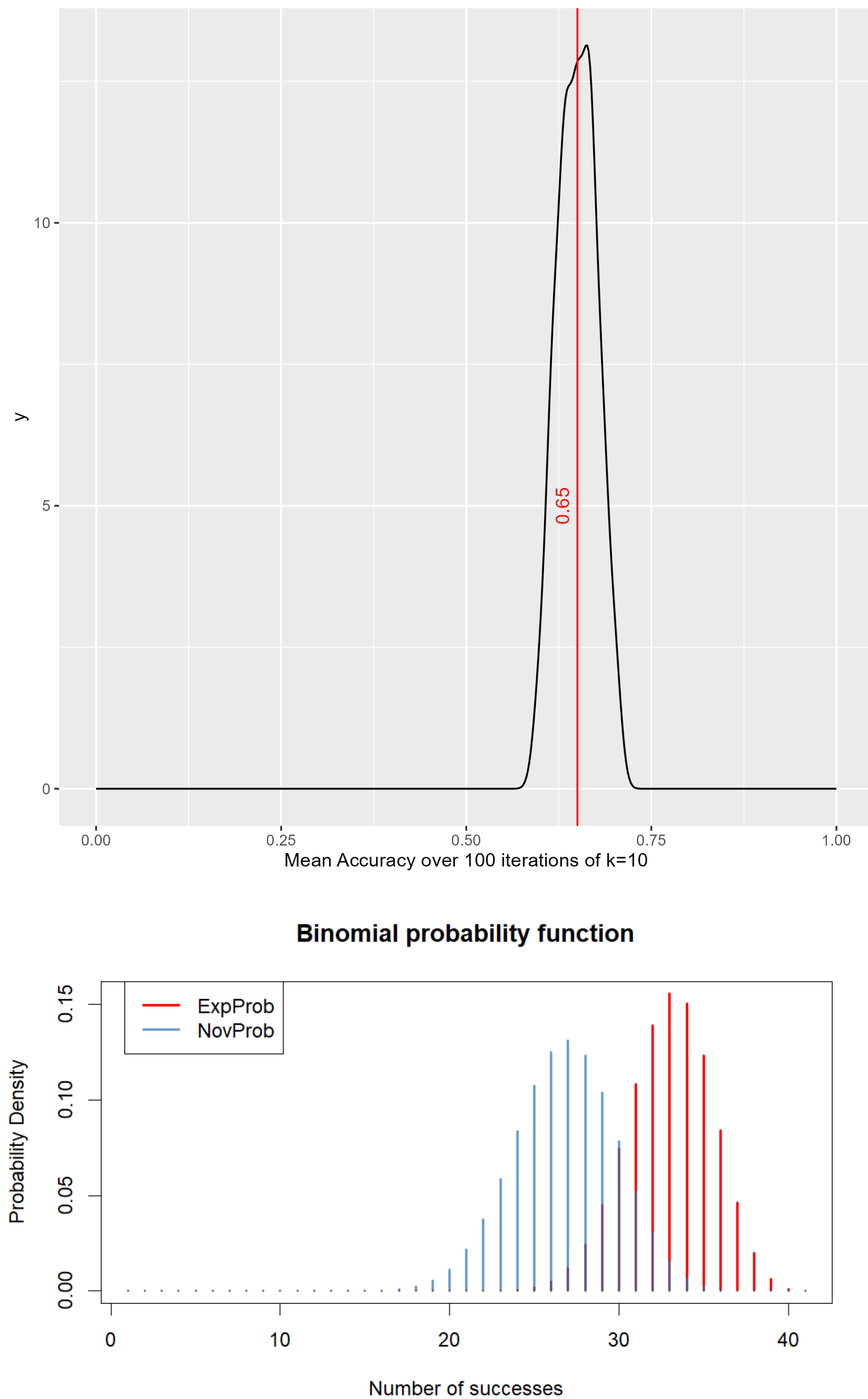


Figure 4.7: Mean prediction accuracy for the ridge regression model (above) and the binomial probability distributions for scores in each driving experiences group, demonstrating the probability of a single participant's number of success for the two driving groups (below).

4.4 Chapter 4 Discussion

Results from the validation study presented in this chapter indicate that the Hazard Prediction clips that form this test are able to successfully discriminate between Novice and Experienced drivers, demonstrated by their ability to predict the hazards correctly at a significantly higher rate. Furthermore, clip-by-clip analysis demonstrated that it was possible to predict whether a person taking the test was an Experienced or Novice driver based on their scores, with greater success than chance. This indicates that this version of the Hazard Prediction test is able to discriminate between the two groups of drivers and thus can be used for evaluating driver's hazard prediction ability, a vital part of the higher levels of Endsley (1995a)'s model of Situational Awareness. As such, this test was deemed acceptable to use for subsequent experiments presented in this thesis as a probe of situational awareness through measuring hazard prediction ability.

The benefit of using a Hazard Prediction test is that it allows specific measuring of the higher mental representations a driver has of a road environment (Crundall, 2016). By scoring participants' accuracy on their ability to predict what happens next in each clip, it allows an objective measurement of the mental representation they have of the road scene and of the third level of Endsley (1995a)'s Situational Awareness model: *Projection*. This is a better reflection of the cognitive state of a driver, and their ability to navigate the road safely (Crundall, 2016). The multiple-choice Hazard Prediction test used in this chapter is presented as a tool for researchers to use when attempting to measure driver awareness in a series of naturalistic road scenarios. This test has a variety of practical benefits over the standard Hazard Perception test: 1) It is easy to administer, only requiring a basic computer setup with a monitor and keyboard, rather than a complex simulator, 2) The instructions are simple and do not require extensive training with participants to collect data, and 3) It is easy to score, with the multiple-choice responses being scored as either correct or incorrect.

Looking at the individual clip coefficients (Figure 4.6) highlights the challenges in creating an effective hazard prediction measure. The majority of the clips show positive prediction coefficients, which mean that performance on these clips could be used to discriminate between Experienced and Novice drivers. On the other hand, the clips with negative coefficients stand out. It is worth noting, that a negative coefficient does not necessarily indicate that Novice drivers performed better than Experienced drivers on these clips, but rather it was not possible to distinguish between the two groups based on their performance alone. This can be due to a multitude of factors and variations that arise when measuring human perception empirically; driver cognition is complex and it is not possible to capture the full mental representation a driver has about a road scene from a video-based test.

However, one important consideration is the choice of occlusion point for each clip, when the video stops and the awareness of participants is tested. Crundall (2016) highlighted the challenges in choosing these occlusion points for a hazard prediction clip; it must contain enough information about the upcoming hazard, its *pre-cursor*, to successfully discriminate between experienced and novice drivers. Indeed, Crundall (2016) found that later occlusion times favoured novice drivers more than earlier occlusions. Here, the clips with the lower coefficients (see Table B.1) show only small differences between the two groups, with very high scores for both on Clip 30 (Experienced 89%, Novice 92%) while Clip 34 shows very low scores (Experienced 33%, Novice 49%). Choosing the cutoff point for each clip provides the main challenge. Though the guidelines set out by Goodge et al. (2021) provide a methodology, validating the test by comparing scores between Experienced and Novice drivers ensures that the clips used can empirically distinguish between these two groups. The results presented here demonstrate that it is possible to use scores on this Hazard Prediction test to discriminate between Experienced and Novice driver groups, indicating that Experienced drivers were able to use their driving experience to more accurately predict what hazard was about to happen next based on the precursors evident in the clips. As such, this version of the Hazard Prediction test is suitable to use as an assessment tool for probing the situational awareness of drivers and their ability to maintain awareness of a road scene.

There are some limitations to this experimental paradigm, nominally that data was collected online which meant relying on the screening procedures from the Prolific online platform (Prolific, 2024). This highlights some issues, such as two drivers in the Novice group reporting a higher number of years experience than the average of their cohort. After following up, these participants did confirm they had not yet qualified for their driving license and so were still classified as Novice drivers. Additionally, during the analysis, models including years of experience were fitted and years of driving experience was not found to be a significant factor in the Novice driver condition, and so these participants were retained. Another potential issue with presenting this test as an online experiment was participants viewed the test on their own devices. While performing the test on mobile devices was prevented, the responses to individual clips may have been affected by variation in devices they were displayed on (i.e., screen brightness, viewing distance, etc.), which were impossible to control. More broadly, the viewpoint of the clips in this test, similar to the clips used in the Hazard Perception test, is somewhat restricted to forward view of the driver. Other iterations of the Hazard Prediction test make use of wing and rear view mirror information (Crundall and Kroll, 2018; Kroll and Crundall, 2019). Further to this, 360 immersive video clips have been developed which emulate the full view afforded to a driver of a vehicle (Goodge et al., 2021).

However, the goal of this version of the test was to create a validated and easy to administer tool that does not require significant investment in equipment to display; a monitor and keyboard are enough to administer this hazard prediction test. While 360 views may offer more ecologically valid hazard clips, the version presented here benefits from a low barrier to implementation.

Finally, the clips that make up this test were all recorded in a limited geographical area. Hazards that make up the clips in this test are naturalistic and were encountered on the road around the Greater Glasgow area. Naturally, this means certain less common hazards in this area could be underrepresented in this test. For instance, there are fewer rural hazard clips as there were fewer hazardous events encountered during drives in these areas. Conversely, there are multiple hazards from busy urban high street areas, due to the complex nature of these road scenes. It could be argued that this is indicative of the demand of these types of driving, but this test does not allow for specific comparison between (sub)urban driving and rural driving ability for instance. Furthermore, all the clips were filmed in a Left-Hand Drive - (LHD) country, and may appear unfamiliar to drivers in a Right-Hand Drive - (RHD) country. However, previous research from Lim et al. (2014) and Ventsislavova et al. (2019) have shown that Hazard Prediction test benefit from cultural agnosticism, and does not discriminate between drivers from LHD or RHD countries, but rather their driving expertise. Ventsislavova Petrova (2019) observed distinct differences between Chinese, Spanish and UK drivers on a Hazard Perception test, but not on a Hazard Prediction version. For instance, Chinese drivers responded less often to hazards in the perception test than UK drivers, which the authors attribute to the difference in cultural hazard thresholds: drivers in China experience a greater number of hazards during their everyday drives, potentially desensitising their definition of a hazard compared to UK drivers (Ventsislavova et al., 2019). While operationally a challenge when changing from a LHD vehicle to a RHD one, this does not appear to impact Hazard Prediction ability.

Though this version of the Hazard Prediction test was not specifically evaluated in a cultural context, when country was included in the analysis procedure it did not provide any greater explanation of the data. Additionally, Goodge et al. (2024) found no differences on performance of this test presented here between drivers who had gained their license in a range of countries when modelling their hazard prediction performance when performing a range of NDRTs in different presentations (described further in Chapter 5). This indicates that, despite geographical limitations in creating the clips, the Hazard Prediction test used here is also not sensitive to cultural differences in driving, concordant with previous literature (Ventsislavova Petrova, 2019; Ventsislavova et al., 2019). As such, it can be used to assess the situational awareness of drivers, regardless of their driving background.

4.5 Conclusion

This chapter presented a version of the Hazard Prediction test which was successful at discriminating experienced and novice drivers. The Hazard Prediction test has benefits over other methods of measuring driver awareness, as it probes the advanced states of situational awareness of projection (the ability to predict what happens next in a given scenario). This is useful for measuring a driver's overall continual awareness of a road scene, rather than using response time measures. This version of the Hazard Prediction test was used in Chapters 5 and 6 to measure situational awareness and address **RQ2** and **RQ3** of the thesis (section 1.3).

Chapter 5

Evaluating the Effects of Heads-Up versus Heads-Down Displays for Non-Driving Activity on Driver Hazard Prediction

Figure 5.1: Examples of the Augmented Reality cues from Experiment 3 (left - gem popping game presented in AR in front of road environment) and Experiment 4 (right - keypad task).

This chapter presents two experiments comparing driver performance on a hazard prediction while performing a distracting task presented as either a heads-up display or a heads-down display. The results show that a heads-up presentation by itself does not aid hazard prediction. Including an attentional cue as part of the heads-up display aided situational awareness over a heads-down presentation, but this was dependant on the workload demands of the NDRT. The results address RQ1 and RQ2 of the overall thesis, as well as providing novel insights for designers of in-car systems about how to present NDRTs to aid driver situational awareness in partially automated vehicles.

5.1 General Introduction

When a vehicle is in autonomous mode, the driver can engage in Non-Driving Related Tasks - (NDRTs), such as reading, playing games or using a smartphone. Taking away the responsibility for driving and freeing up time for NDRTs is one of the main motivations for purchasing an Automated Vehicle - (AV) (Le Vine et al., 2015; Panagiotopoulos and Dimitrakopoulos, 2018). However, current automated vehicles still require driver supervision as a failsafe; the driver must be ready to take over if the vehicle can no longer drive itself in case a Take-Over Request (TOR) is issued. Prolonged supervision is not a task that humans are predisposed to (Bainbridge, 1983), and studies have shown how fatigue and low mental workload are present at higher automation levels (Figalová et al., 2023). The current assumption is that a timely alert for a TOR is sufficient for drivers to change their role from a passive supervisor to an active controller. Previous research has shown that if drivers are not fully situationally aware of the road, their judgement is impeded (Endsley and Kiris, 1995; Morales-Alvarez et al., 2020). Furthermore, it takes time for awareness to be acquired, which may not be available in the case of an urgent TOR (Gold et al., 2016). The lack of motivation to maintain vigilance during an automated drive, paired with a TOR which may not provide adequate information for the driver to make a decision, has dangerous implications for road safety.

Engagement with NDRTs can help reduce fatigue and benefit attention during prolonged supervision (McKerral et al., 2023), but little research has investigated the situational awareness state of the driver supervising an AR whilst engaged with an NDRT. Having awareness of the vehicle and its environment are key for a successful TOR (Morales-Alvarez et al., 2020), which is reduced if the driver is engaged with an NDRT (Janssen et al., 2019). The challenge is how to keep drivers situationally aware of their driving environment without sacrificing the benefits of automation.

Augmented Reality - (AR), where virtual objects are superimposed onto the real world (Azuma, 1997), has been suggested as one solution to keeping drivers in the loop through an AR Heads-Up Display - (HUD); displaying virtual content at eye level with the road (Riegler et al., 2019a). Compared to a Heads-Down Display - (HDD), an AR HUD allows the driver to interact with non-driving related content overlaid on top of the view out of the front of the car, potentially allowing them to monitor the road and be ready if a TOR occurs. Much research has demonstrated the benefits of HUDs on driver performance (Crawford and Neal, 2006; Liu and Wen, 2004a). However, few have investigated how interacting with an NDRT through a HUD impacts the underlying perception of the driving situation and road hazards, necessary for an effective TOR.

Many studies demonstrate the 'Look But Fail To See' phenomenon, where humans perceive but fail to adequately process information presented to them (Simons and Chabris, 1999; Hills, 1980; Wolfe et al., 2022). This also affects the perception of a driving scene, including drivers failing to see pedestrians, cyclists and motorcycles despite fixating on them (Crundall et al., 2012; Kaya et al., 2021; Beanland et al., 2014). This is key for ascertaining if drivers can maintain situational awareness whilst engaged with an NDRT in an AV context. Just because a driver fixates on a hazard does not mean they are attending to it. AR is a potentially useful method of presentation for these NDRTs, but it is still unclear whether the heads-up view would facilitate situational awareness or allow them to perform the NDRT effectively. Previous work highlights how including driving-related cues in an informational AR HUD can aid driver situational awareness (Rusch et al., 2013; de Oliveira Faria et al., 2021). If displaying NDRTs in an AR HUD by itself does not aid situational awareness, could including an additional attentional cue combat the 'Look But Fail To See' effect?

This chapter presents two experiments which investigate how the Situational Awareness - (SA) of drivers is affected by engaging with an NDRT. Presentation method for NDRTs was compared between AR HUDs (both with and without an attentional cue) and both AR and traditional HDDs (the current format used for vehicle info-tainment displays), as well as to a Control condition where participants only focused on the driving task. The ability to predict hazards (a key component of SA), confidence and subjective attention were measured, as well as perceived workload and performance on the NDRT. Results showed that participants were able to maintain some awareness whilst engaged with an NDRT in all presentation conditions, but always worse than when solely watching the road. Including an attentional cue in the AR HUD increased awareness compared to the HDD condition, but only when the NDRT was less demanding. This suggests that current heads-down presentations of in-car NDRTs are not suitable for keeping drivers situationally aware of their environment. New presentation methods for NDRTs are necessary to aid driver awareness of the road. This chapter discusses the implications of this and how future in-car displays should be designed to facilitate driver awareness as well as the suitability of AR as a presentation method.

5.2 Related Work

Driver attention in automated vehicles

Driving is a complex cognitive task which requires quick appraisal and decision-making skills to take safe and appropriate actions. Many factors must be taken into account to inform these decisions, which can only be done when drivers are fully aware of the driving environment (Walker et al., 2009). The advent of more sophisticated AVs allows drivers to hand over control of driving tasks, such as speed control, lane discipline and hazard perception, to the vehicle. This changes the role of the human driver to more of that of a passenger, and allows them to engage with NDRTs safely (McGill et al., 2017; Kun et al., 2018).

It has been shown that driver ability to maintain supervision of automated tasks is limited (Head and Helton, 2014). However, Level 3 AVs currently available to consumers still require a human operator to maintain vehicle supervision as a fail-safe (Society of Automotive Engineers, 2018). To alleviate this, AVs employ alerts to indicate when the driver should take control of the vehicle, known as a Takeover Request (TOR). A significant amount of work has been invested in researching and designing TORs that can effectively alert the driver (Peck et al., 2015; Politis et al., 2014b, 2017; Salubre and Nathan-Roberts, 2021). However, being alerted to a hazardous event is different to appraising it. Failure to perceive hazards reveals a fundamental failure in what Endsley (1995a) describes as having *'Situational Awareness'* (Endsley, 1995a): the ability to 1) perceive the environment, 2) comprehend what is occurring and then 3) predict what might be about to happen based on prior knowledge. If a driver does not perceive a motorcycle or pedestrian, they will not be able to predict their actions, much less act safely should they be required to. Further to this, if drivers struggle to maintain situational awareness whilst fully in control of a vehicle, this is worsened when supervising an AV where there is no requirement or motivation to be engaged in the driving task (Endsley and Kiris, 1995; Pipkorn et al., 2022; Endsley, 2019). Studies investigating sustained attention on automated processes indicate a marked drop in cognitive performance as the time spent monitoring increases (Head and Helton, 2014; Esterman and Rothlein, 2019), which suggests that drivers are likely to neglect these important supervision tasks.

Driver gaze behaviour in automated vehicles

Driver's eye movements have long been used to estimate attention on the road (Underwood et al., 2003; Crundall and Underwood, 2011) and the relationship between fixation patterns and driver attention is a well-established finding (see Arias-Portela et al. (2024) for review).

However, fixation patterns between drivers of manual vehicles differ to those supervising in an automated vehicle. Mackenzie and Harris (2015) found that driver's eye patterns differ between passively watching a drive and actively controlling a vehicle, with fixations located closer to the front of the vehicle during an active drive and greater horizontal scanning in a passive drive. Subsequent research corroborated these findings and suggested that drivers supervising an AV exhibit more 'Lookahead' fixations (Mole et al., 2021) further down the road to anticipate whether they should need to take back control. More broadly, eye movement distributions have been found to be distinct between manual and automated driving (Tang and Guo, 2019) and can used to predict a driver's reaction time (Wu et al., 2021), with larger saccades associated with faster reactions to an alert. Drivers can even be classified as low or high risk, depending on their gaze behaviour (Zeeb et al., 2015). These studies indicate that eye movements are a useful measure for determining how aware of the road a driver is. As such, it is the basis for some driver monitoring systems used to evaluate whether a driver is paying attention to the road or not (Vicente et al., 2015), as there are notable differences between a distracted driver's eye movements and one focusing on the road (Sodhi et al., 2002; Le et al., 2020).

However, the purpose of automated driving is to allow the driver to engage in other non-driving-related activities. This typically necessitates them to take their eyes off the road to engage with an NDRT, which has typically been shown to be detrimental to driving performance (Oviedo-Trespalacios et al., 2016; Yannis et al., 2014), and therefore safety. Previous research looking at the effect of mobile phones found that having a conversation impaired recognition of stimuli presented at fixation (Strayer et al., 2003), and disruption occurs even during hands free interaction (Desmet and Diependaele, 2019). Merat et al. (2014) found that road-relevant eye movements are disrupted by an unforeseen takeover request and could take up to 40 seconds to recover. Combined with the fact that engagement within a demanding NDRT leads to longer reaction times to a takeover request (Wintersberger et al., 2018b; Müller et al., 2021), it is important to evaluate whether a driver supervising an AV is able to regain full attention of the road after engaging with a distracting NDRT.

Keeping drivers in the loop

One suggestion for providing information to the driver is to use a HUD, where information is presented at the driver's eye level. Following research demonstrating their benefits in the aviation industry (Ingman, 2005), HUDs are now appearing in cars. They provide easier access to relevant information compared to a traditional HDD presented via an instrument cluster or a centre console, which require drivers to take their eyes off the road.

Previous work exploring HUDs in cars suggests that driving performance was less impaired and preferred over traditional cockpit displays (Medenica et al., 2011; Smith et al., 2015; Jose et al., 2016). The use of HUDs as a driver awareness aid, e.g., a crash warning system, has been shown to help reduce mental workload (Schömig et al., 2018), reduce reaction times in manual driving (Kim et al., 2013; Wintersberger et al., 2018b) and improve performance on both high and low demand NDRTs (Li et al., 2020) in manual driving. However, current HUDs are limited to small, unintrusive displays showing static information such as speed or navigational aids. There are important considerations for designing HUDs to display more content to drivers in a way that is not distracting from the driving task. In particular, factors such as the visual complexity of the driving scene and mental workload affect a driver's eye movements and visual scanning patterns (Crundall and Underwood, 1998; Kapitaniak et al., 2015). A busy and attention-capturing HUD is more likely to distract attention than assist awareness of the road (Lee et al., 2020; Kim and Gabbard, 2022), yet there is a lot of information that drivers need to be kept informed of, especially if supervising an AV. A balance needs to be struck between presenting information to the driver that keeps their eyes on the road, but is not so overloading that it intrudes on the driving task.

Using Augmented Reality to display information to drivers

Augmented Reality - (AR), where virtual images are superimposed onto the real world (Azuma, 1997) has become a popular means of displaying information to drivers. An AR HUD is distinguished from a conventional HUD as it allows more detailed information to be displayed in a dynamic fashion, such as highlighting specific objects on the road (Gabbard et al., 2014; Karatas et al., 2020; Riegler et al., 2019a). AR HUDs with more dynamic visual cues have also been shown to aid driving performance. Jing et al. (2022) found that AR HUDs were able to reduce distraction when focusing on dangerous driving scenarios, and Bark et al. (2014) showed that a navigational AR HUD aided turn decisions. Lindemann et al. (2018) found that an AR HUD showing a variety of driving-related information such as threat markers and oncoming traffic indicators improved situational awareness of drivers. This finding was echoed by Karatas et al. (2020) who showed that an AR HUD highlighting hazards led to quicker recognition compared to a traditional HUD. Rusch et al. (2013) showed that specifically directing attention with AR cues increased detection rates of pedestrians and warning signs. Similarly, Wang et al. (2022b) found that an AR HUD which did not highlight pedestrians led to reduced rates of recognition, not evident when the same pedestrian was pointed out to drivers.

Nachiappan et al. (2021) found that, despite perceptions of increased workload, a letter recognition NDRT presented via an AR HUD increased driving performance during monotonous manual drives; though this came with increased attention to the AR HUD and not the road. In an AV context, de Oliveira Faria et al. (2021) found that AR cues helped improve driver behaviour after a TOR and reduced driver-initiated TORs. This research suggests that a HUD presentation benefits driver attention by keeping their eyes on the road and reducing the time it takes for them to react to any dangers on the road. To further enhance the chance of a driver noticing dangers on the road, AR can be used to dynamically highlight objects in the road scene. Cueing in AR headsets has been shown to benefit target detection (Yeh and Wickens, 1999), visual search at a distance (Warden et al., 2022) and a virtual-navigation task (Stefanucci et al., 2022). In-car AR also has shown benefits, where highlighting pedestrians or other road targets through an AR HUD appears to successfully draw driver attention (Kim and Gabbard, 2022; Rusch et al., 2013; Karatas et al., 2020; Wang et al., 2017b), although this can come at the expense of distracting drivers from other road objects (Wang et al., 2022b).

Does Fixation mean Attention?

However, the evidence supporting the use of AR HUDs typically evaluates displaying *driving-related* information to drivers of automated vehicles is beneficial for their awareness. These studies typically measure driver performance when supervising the driving task in a L1 or L2 setting. The case is not so clear when presenting an intentionally distracting NDRT that draws attention away from the road. This is the likely application of AR HUDs in AVs (Schroeter et al., 2014; Gabbard et al., 2019; Li et al., 2020), incorporating NDRTs into AR displays to encourage attention to important aspects of the road (Schroeter and Steinberger, 2016; Schroeter et al., 2014). This concept has been demonstrated for passengers (Togwell et al., 2022; McGill et al., 2017), and Muguro et al. (2021) found that interacting with a gamified AR HUD reduced reaction time to popup traffic events. Steinberger et al. (2017) also found that a gamified coasting challenge in AR was shown to reduce boredom on long simulated drives. This previous work indicates that AR HUDs could be beneficial in providing information to drivers to improve both their driving ability and their SA. Once again however, these studies all present driving-related information to assist driver attention and their awareness of the road. A meta-review by Niu et al. (2024) highlights how warnings presented in a HUD decrease reaction time compared to HDDs. They note, however, that few studies investigate the specific use of *AR* HUDs and their potential benefits, in particular the uncertainty around how reaction to warnings and hazards are affected by engagement with an NDRT.

Despite this, the UK & Scottish Law commission state in a joint report that NDRTs are permissible in AVs if they *"do not prevent the driver from responding to demands from the automated driving system"* The resolution states that the user should be *"ready and able to take control"* and *"maintains the capabilities necessary to fulfil their respective duties."* (UK and Scottish Law Commissions, 2020). It is unclear whether engaging with an NDRT when supervising an AV affects a driver's ability to maintain awareness of the road.

Li et al. (2020) demonstrated that presenting an NDRT in a HUD versus a mobile phone does help drivers keep their gaze on the road ahead. However, this relies on the assumption that fixating on an object leads to greater processing. While typically true for undistracted viewing (Velichkovsky et al., 2000), this is not always the case when the person fixating is distracted by another task. Simons and Chabris (1999) famously demonstrated this effect of *"inattentional blindness"*, when participants asked to attend to basketball players in white shirts and ignore those in black shirts failed to perceive a person in a gorilla costume walk across the scene. Follow-up eye tracking research found that fixation was not a predictor of recognition (Pappas et al., 2005; Memmert, 2006; Gelderblom and Menge, 2018), with participants who 'saw' the gorilla not noticing its presence when asked about it. Similar research has found comparable effects with participants not recalling information on advertisement banners (Gelderblom and Menge, 2018; Resnick and Albert, 2014), expert radiologists missing a gorilla image superimposed on a lung CT scan (Drew et al., 2013), and participants not noticing a sudden image of a spider appearing during a visual discrimination task (Wiemer et al., 2013). The effect persists even when stimuli are primed beforehand (Oktay and Cangöz, 2018). This is despite the fact that, in all of these studies, participants were shown to actually be fixating on the target stimuli.

Taken into an automated driving context, this has some serious implications for the safety of using a HUD to display non-driving content. A driver distracted by an NDRT is asked to take back control of an AV. If this NDRT is presented at eye-level in a HUD, are they able to maintain awareness of the road throughout and seamlessly take control of the vehicle safely? Alternatively, does the distracting nature of an NDRT mean that, despite fixating on the road, a HUD presentation provides no benefit and our driver finds themselves thrust into a dangerous situation with little or no preparation, suddenly in control of the vehicle? It is not enough to simply measure the gaze behaviour of a distracted driver; it is also important to include a measure of their situational awareness and understanding of the road scene.

Eye-tracking in AR

Traditional eye tracking measures focus on viewing the screen with 2D-presented content. Taking it into the real world with portable eye tracking glasses still only provides a 2D canvas with x,y coordinates of what a person is viewing. With the advent of more sophisticated AR and VR headsets, it is possible to collect eye tracking from participants viewing content in a 3-dimensional environment. Kapp et al. (2021) put forward their Augmented Reality Eye Tracking Toolkit - (ARETT) which allows eye tracking data to be collected in Microsoft's HoloLens 2 AR headset. This allows the implementation of eye-tracking into a Unity scene and the collection of eye-tracking data in a 3D environment. By setting up an AR environment with Areas of Interest - (AOIs) defined in 3D space, ARETT (Kapp et al., 2021) allows the collection of eye movement data in a full 3D environment. In the context of presenting NDRTs via an AR HUD, this allows the measurement of whether a driver is fixating on the road scene in the distance, or the AR display which is presented in the near focus. In a 2D analysis, a driver's fixation may land square on a hazard, but they are actually focused on the display which has been overlaid on top of it. If AR HUDs are to be used for presenting any information to drivers of automated vehicles, it is useful to measure whether a driver's fixations are actually on the road environment.

Summary and Research Questions

Whilst previous studies have established the benefits of presenting information to drivers via an AR HUD, the effect of presenting an NDRT in this way on situational awareness is still not clear. Questions remain over whether presenting an NDRT in a HUD benefits situational awareness, or serves as a distraction from the driving task, as demonstrated by the Look But Fail To See phenomenon. The experiments presented here measure performance on a hazard prediction task compared between different NDRT presentation methods of an eye-tracking task (Experiment 3) and a keypad input task (Experiment 4). This chapter sets out to address the following research questions:

- **Ch5 RQ1** Can drivers maintain situational awareness when they are engaged with a NDRT?
- **Ch5 RQ2** Does presenting a NDRT via a HUD have benefits for situational awareness over a traditional HDD?
- **Ch5 RQ3** Does including attentional cues to direct attention to the road in the design of the NDRT aid situational awareness?

5.3 Experiment 3

Comparing the effects of AR HUD vs HDD presentations of an NDRT on Hazard Prediction Ability

An experiment was conducted to compare driver situational awareness, as measured through their Hazard Prediction ability, whilst engaged with an NDRT presented either via a HUD or a HDD. The experimental design was approved through the College of Science and Engineering Research Ethics committee (Application number #00220199). The following section reports the methods, procedure and results from this experiment.

Design

The experiment was designed to measure how Hazard Prediction ability was affected by performing an NDRT across different presentation methods. A repeated measures experimental design was employed, with Hazard Prediction score and subjective confidence rating as dependent variables. The independent variable was Presentation Method with five levels: Baseline (no NDRT), AR HUD, Cued AR HUD, AR HDD and Tablet HDD.

The Hazard Prediction test scores were used to answer RQ1. If scores were above the chance level of 25%, this indicated that the participants were able to predict what happened next in the video clips and supposedly would be able to take over control of an AV safely. Differences in the scores between each of the presentation methods would answer RQ2. If scores in the HUD conditions are higher than in the HDD which takes attention down off the road, it suggests that this eyes-on-road presentation method provides some benefit to situational awareness. To answer RQ3 and measure a lack of awareness caused by the 'Look But Fail To See effect, an AR HUD condition which used a specific attentional cue to direct attention towards the hazard was included. This was to evaluate whether an AR HUD by itself is beneficial for situational awareness or if it requires specific design considerations to do so. Feedback about how demanding each of the presentation conditions were and how confident participants felt in their answers provided insight into the workload caused by engaging with an NDRT for each presentation method.

Participants

A power analysis was conducted using the *pwr* package (Champely et al., 2016) in R Studio, which determined that for a moderate effect size of 0.8, 22 participants were required for the repeated measures design. Twenty four participants *(Mean Age = 33.1 years, SD = 9.4, 11 Female)* were recruited via online forums and around the University of Glasgow Computer Science and Psychology departments.

All had normal or corrected to normal eyesight and had held a driving licence for at least 2 years. Since previous research has shown the Hazard Prediction test to be culturally agnostic (Ventsislavova et al., 2019), recruitment was not limited to drivers from the UK (10 UK, 2 Germany, 2 Greece, 1 Denmark, 1 France, 1 Italy, 1 Taiwan, 1 Spain, 1 Philippines, 1 Malaysia, 1 Indonesia, 1 Bulgaria and 1 held licenses from both Saudi Arabia & New Zealand).

The average total driving experience was 13.7 years *(min = 2, max = 41, SD = 8.04)*, and the average UK driving experience for non-UK license holders was 2.04 years *(min = 0, max = 6, SD = 2.04)*. Eighteen people reported they had experience of driving in the Glasgow area where hazard clips were filmed, with an average of 3.7 years *(min = 0, max = 26, SD = 6.16)*. Nine reported having used an AR headset before, seven reported using mobile AR and six reported never having used AR. One participant reported never having heard of AR.

Materials

Hazard Prediction Test

The Hazard Prediction test (What Happens Next? - WHN) as set out in Chapter 4 was used as a measure of SA. Participants were presented with the 40 hazard clips whilst sitting in a driving simulator. Since it was not possible to capture unobstructed footage from the side windows of the vehicle, the side monitors were turned off to avoid distracting participants and only the forward view out of the windscreen was presented, akin to the current version of the Hazard Perception test (Driving and Vehicle Standards Agency, 2023). The experiment was built using PsychoPy v2021.2.3 (Peirce et al., 2019) and displayed on a Samsung LC32R500FHRXXU 32in curved monitor approximately 1m from their face. A multiple choice list of 4 potential scenarios was presented, where they selected one using a button on a Logitech G29 steering wheel (see Figure 5.2).

Design of the NDRTs

Four presentation conditions were designed to display the NDRTs: AR HUD, Cued AR HUD, AR HDD and Tablet HDD. The AR tasks were developed in Unity (version 20203.26f1) using the Mixed Reality Toolkit (MRTK - version 2.7.2) and were presented using the HoloLens 2 Augmented Reality headset. Guidelines for placing and displaying Mixed Reality content from Microsoft (2023) were followed for the placement, size and opacity of images, where it did not interfere with the design of the experiment. Images were displayed at the same distance from the user as the real life screen in order to reduce potential discomfort caused by shifting focus between near and far objects.

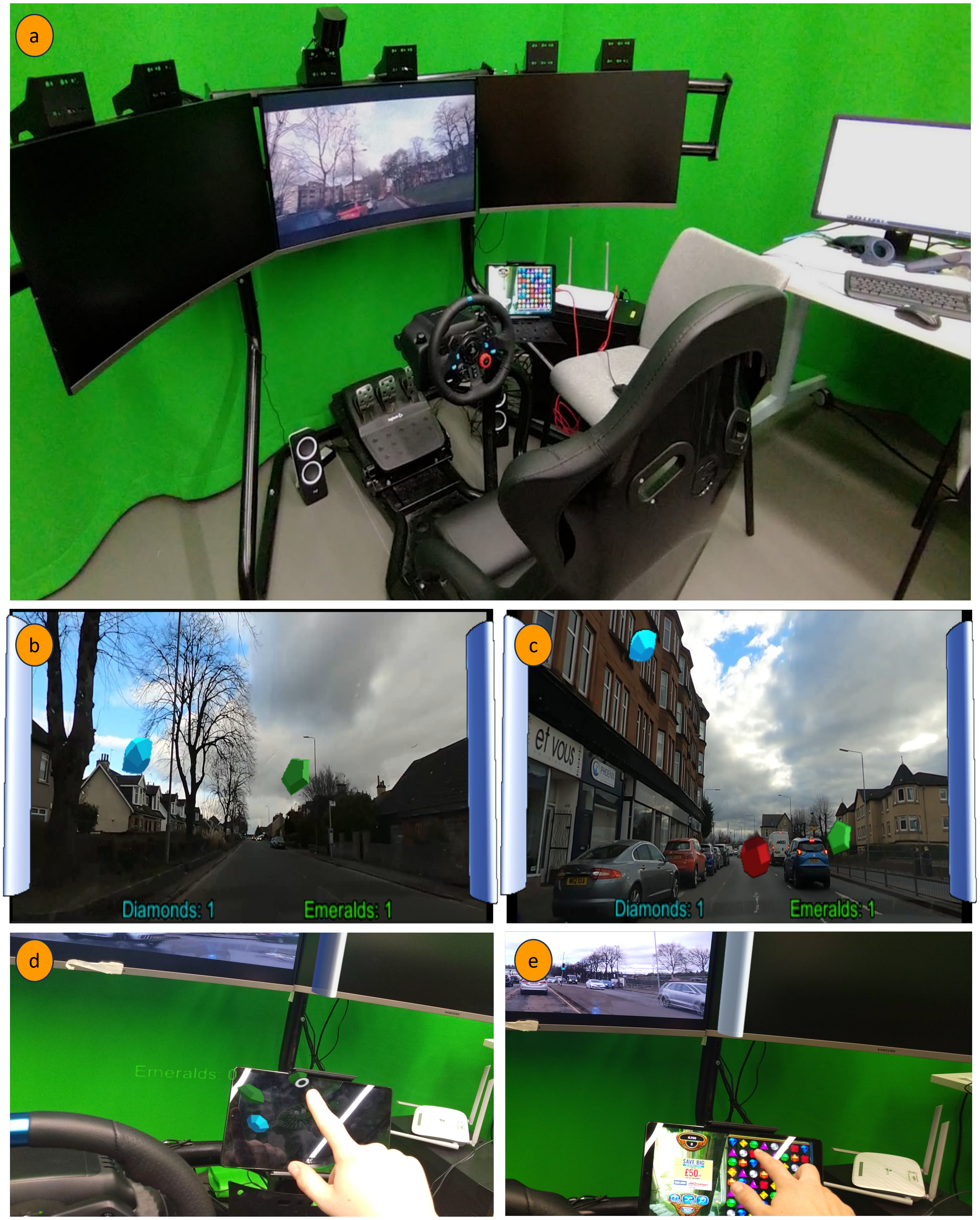


Figure 5.2: Experimental setup, with a) the What Happens Next - (WHN) Hazard Prediction task on the centre screen and the Tablet displaying the NDRT mounted on the simulator rig, with representations of the b) AR HUD c) Cued AR HUD d) AR HDD and e) Tablet HDD conditions.

The same AR game presented in Chapter 3 was used, with a design similar to that used by Radlmayr et al. (2018) and gamified features added, such as a visible score count. Coloured gems would appear in the 3D space in front of the monitor displayed at random intervals. Participants were asked to 'pop' all the gems they could as quickly as possible by looking at them. The gems lasted for between 1.5 and 2 seconds, and participants received a point for each gem they popped. Performance was measured by counting the number of gems popped compared to the number of gems spawned to get their accuracy level. The gems stopped spawning when the Hazard Prediction prompt appeared. An eye-tracking task was selected as it did not require the hands, which could rest on the steering wheel and allowed the hazard clips to be easily visible. This task was world-locked to the monitor to emulate a windscreen display. For the AR HUD condition, the gems appeared randomly in front of the screen. This condition was included to measure whether a HUD presentation would provide any benefit over a HDD. In the Cued AR HUD condition, a red gem would appear as an attentional cue in the area below where the hazard on the screen would appear 4 seconds before its onset, based on findings of average takeover request times from Eriksson and Stanton (2017). This condition was used to examine the effect of the Look-but-Fail-to-See phenomenon, comparing performance with the AR HUD conditions.

In the AR HDD condition, the same game task as described above was displayed down below the level of the screen to bring participant's attention away from the road. They used mid-air touch gestures, rather than gaze tracking, detected by the HoloLens 2, to interact with the gems and score points. This was designed to emulate interacting with a central console that is currently being suggested for NDRTs, but still within the AR domain. A Samsung Galaxy tablet was mounted onto the driving simulator for the Tablet HDD condition (see Figure 5.2). This tablet condition was chosen to represent the type of NDRT drivers are likely to engage with, as well as what is currently legal in the UK. The mobile game Bejeweled (Electronic Arts Games, 2021) was presented, which requires players to align gems into a row of three to clear the board and gain points. Whilst the demand of this task and the input methods are different from the AR NDRTs, it was included as a realistic alternative to the other NDRT conditions. Though specific comparisons between this task and the others cannot be made, it provides a broad indication of the differences between an AR NDRT and the touchscreen based one that are currently available in cars, such as recent Tesla models (Tesla, 2023) and BMWs iDrive (Hawkins, 2021).

Dependent variable	Scale	Timepoint
Hazard Prediction scores	Correct / Incorrect	After each clip
Confidence rating	0-100 scale	After each clip
Subjective Attention rating	0-100 scale	End of Block
NASA TLX	6 item 0-100	End of Block
SSQ	18 item 0-3	End of Block
NDRT Performance	Proportion of gems hit	End of Block

Table 5.1: Dependent variables measured and what timepoint in the experiment they were collected. The total correct scores of the 8 Hazard Prediction clips per condition, as well as the average of each of the other measures between each condition were compared.

Procedure

After consenting to take part, participants provided demographic and driving experience information via the online questionnaire platform Qualtrics. They were then shown an example WHN clip to practice giving their responses, as well as given a practice interaction with the headset. Participants saw the 40 WHN clips in 5 blocks (one for each presentation condition and with 8 WHN clips making up one block). First, they performed the Hazard Perception task without any NDRT as a baseline, and then proceeded through the four counterbalanced NDRT task conditions with 24 iterations, meaning no two participants had the same order of NDRT task condition. After watching each clip, participants were asked to predict what happens next from the list of multiple-choice answers. They were also asked to rate their confidence in their answer on a 0-100 scale. The NASA Task Load Index - (TLX) (Hart and Staveland, 1988) was administered via Qualtrics after each condition to assess perceived workload. Finally, participants were asked to rate their attention to the driving task on a 0-100 scale at the end of each condition (see Table 5.1 for a list of measures and Figure 5.3 for a diagram of the procedure). The experiment took around 60 minutes to complete. On top of the £10 compensation for taking part, participants were told they could win an extra £5 reward if they performed the best in both the NDRT and the WHN task out of all other participants, to incentivise attention and performance on both tasks.

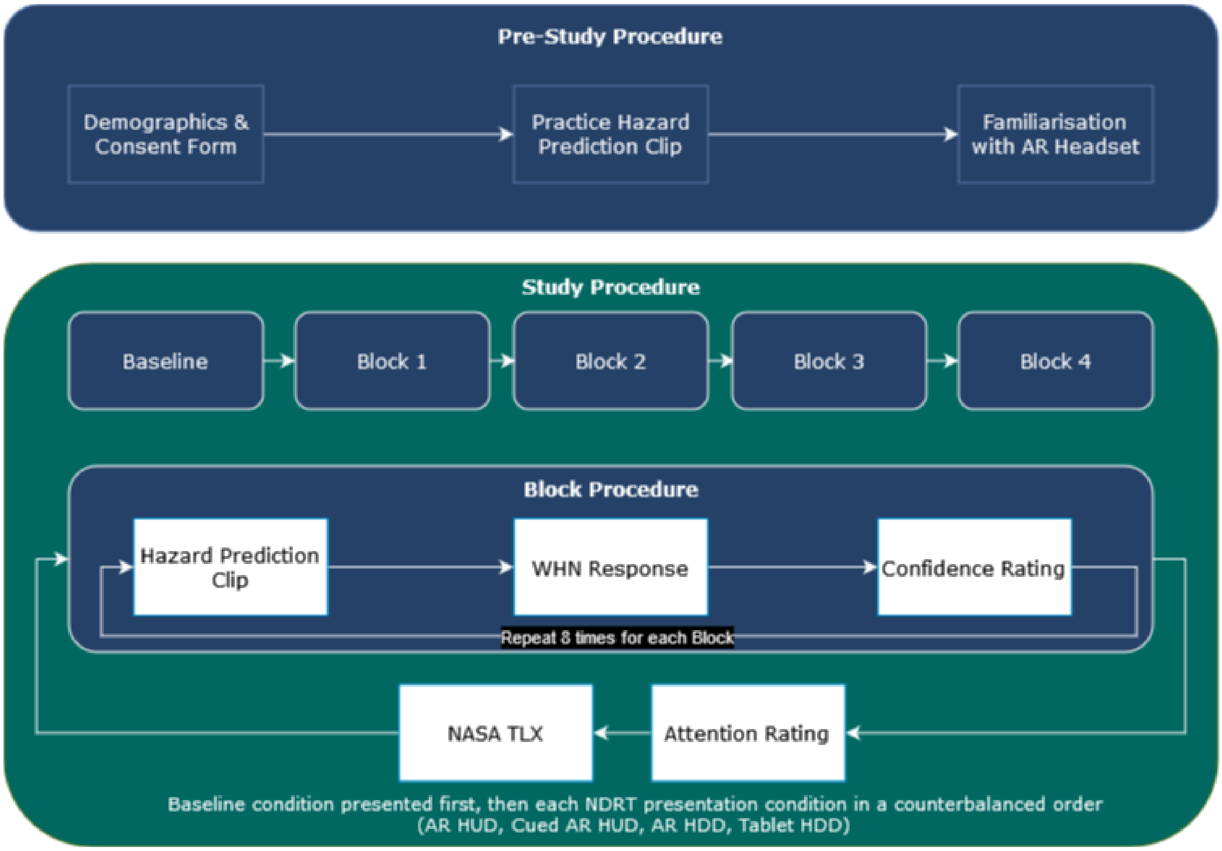


Figure 5.3: Flowchart showing the procedure of Experiment 3.

5.4 Experiment 3 Results

The results were analysed using R Studio (2022.07.01, Build 554) using the lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017) and report (Makowski et al., 2023) R packages. Given the hierarchical nature of the data from the repeated measures design, Generalised Linear Mixed Effects - (GLME) models were fitted to the data for Hazard Prediction score and Confidence ratings, estimated using Maximum Likelihood - (ML) and the bobyqa optimizer. Confidence Intervals - (CI) of 95% and p-values were computed using a Wald t-distribution approximation. Mixed design Analysis of Variance - (ANOVA) tests were conducted on the Subjective Attention ratings and NDRT Performance scores as these were not nested data and so a mixed-effects model was not deemed suitable.

This section contains analyses for the AR NDRTs conditions for Hazard Prediction performance, confidence, attention and NASA TLX ratings. Due to differences in task scoring and interaction method, NDRT performance in the Tablet HDD condition cannot be directly compared to the other NDRT conditions. However, as it was meant to act as a real-world comparison to the types of NDRT which are currently legal for automated driving in the UK, it was included in the analyses.

Hazard Prediction scores

Models were fitted estimating the fixed effect of holding a UK license, number of years of driving experience, driving experience in the UK, and local Glasgow driving experience on Hazard Prediction scores. However, following a backward step wise model selection approach where variables are systematically removed from models and their fits compared, none of these models were found to provide significantly greater explanation of the variance compared to the models presented below. Models containing sequentially less significant variables were compared and removed until arriving at the simplest model following the 'keeping it maximal' approach suggested by Barr et al. (2013). As such, all participants were analysed together, regardless of what country they received their driving license from or their driving experience. Average scores for the Hazard Prediction task for each of the 5 Presentation Method conditions (Baseline, AR HUD, Cued AR HUD, AR HDD and Tablet HDD) were compared (see Table 5.2).

A GLME model was fitted to predict the main effects of Presentation Method including hazard clip as a random intercept for each group with the formula:

$$\text{Hazard Prediction Score}_{ij} = \beta_0 + \beta_1 \cdot \text{Condition}_{ij} + u_{0i} + e_{ij}$$ [1]

This model was found to explain significantly greater variance than a null model with Condition fitted as a random effect as well as participant *(Null Model AIC = 1214.1, BIC = 1228.6; Mixed Effects Model AIC = 1181.2,BIC = 1215.3;p < .001)*. The model's total explanatory power was moderate (conditional $R^2 = 0.22$). Within this model, the average Hazard Prediction score for the AR HUD *(Est. = -0.98, 95% CI [-1.47, -0.50], p < .001)*, Cued AR HUD *(Est. = -0.65, 95% CI [-1.15, -0.15], p = .011)*, AR HDD *(Est. = -1.22, 95% CI [-1.71, -0.73], p < .001)* and Tablet HDD *(Est. = -1.37, 95% CI [-1.85, -0.88], p < .001)* conditions were significantly lower than Baseline scores (see Figure 5.4). After refactoring the model to use Cued AR HUD as the intercept, the scores in the AR HDD *(Est. = -0.57, 95% CI [-1.02, -0.12], p = .012)*, and Tablet HDD *(Est = -0.72, 95% CI [-1.17, -0.27], p = .002)* were found to be significantly lower than the Cued AR HUD condition. However, the scores in the AR HUD condition were not significantly different from the Cued AR HUD condition. Refactoring the model with the AR HUD or AR HDD conditions as intercepts produced no significant differences not already accounted for in the models above (see Table 5.3 for a full list of model comparisons).

[1]- Hazard Prediction Score$_{ij}$ is the response variable for the i-th observation in the j-th group. - β_0 is the fixed intercept. - β_1 is the fixed effect coefficient for the Condition variable. - Condition$_{ij}$ is the value of the Condition variable for the i-th observation in the j-th group. - u_{0i} is the random intercept for the i-th group. - e_{ij} represents the residual error term for the i-th observation in the j-th group. The fixed effects are denoted by β coefficients, and the random effects are represented by u terms.

Condition	Prediction Score (P)	Std Err	Lower CI	Upper CI
Baseline	0.82	0.032	0.752	0.879
AR HUD	0.64	0.045	0.545	0.719
Cued AR HUD	0.71	0.041	0.623	0.784
AR HDD	0.58	0.047	0.487	0.668
Tablet HDD	0.54	0.047	0.45	0.634

Table 5.2: Summary statistics for the average probability of a correct Hazard Prediction score for each Presentation method in Experiment 3, as well as the standard error and both lower and upper confidence intervals as reported from the mixed effects model.

Hazard Prediction	Baseline			AR HUD			Cued AR HUD			AR HDD			Tablet HDD		
Model Intercept	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.
Baseline	1.5	0.22		-0.99	0.25	**p <.001*****	-0.65	0.25	**p - 0.011***	1.22	0.25	**p <.001*****	1.37	0.23	**p <.001*****
AR HUD				0.56	0.2		0.34	0.23	p = .142	-0.23	0.22	p = .29	-0.38	0.22	p = .08
Cued AR HUD							0.9	0.2		-0.57	0.23	**p = .012***	-0.72	0.23	**p = .002****
AR HDD										0.3	0.19		-0.15	0.22	p = .49
Tablet HDD													0.17	0.19	

Table 5.3: Model estimates, Standard Error (SE) and p values obtained through Wald's approximation for each of the GLME models for each of the 4 Presentation Methods and Baseline Hazard Prediction scores. Each row corresponds to a model with the named presentation condition as the intercept, the column representing each of the other presentation conditions compared to the intercept. Repeat comparisons are omitted for clarity, but represent the inverse of the estimate presented.

Confidence Ratings

Average scores for the Confidence rating for Hazard Prediction responses in each of the 4 NDRT Presentation Methods (AR HUD, Cued AR HUD, AR HDD and Tablet HDD) were compared to baseline ratings (see Table 5.4). A GLME model was fitted to predict the main effects of Condition including participant and hazard clip as random effects, with the formula:

$$\text{Confidence Rating}_{ij} = \beta_0 + \beta_1 \cdot \text{Condition}_{ij} + u_{0i} + u_{1j} + e_{ij}{}^2$$

The model's total explanatory power was moderate (conditional $R^2 = 0.22$) and was found to explain significantly greater variance than a null model with Condition, Hazard clip and participant fitted as random effects *(Null Model AIC = 1200.6, BIC = 1215.2; Main Model AIC = 1131.7, BIC = 1165.7; p < .001)*. Within this model, the score for the AR HUD *(Est. = -1.13, 95% CI [-1.65, -0.61], p < .001)*, Cued AR HUD *(Est. = -0.59, 95% CI [-1.13, -0.05], p = .031)*, AR HDD *(Est. = -1.86, 95% CI [-2.38, -1.35], p < .001)* and

[2]- Confidence Rating$_{ij}$ is the response variable for the i-th observation for the j-th participant. - β_0 is the fixed intercept. - β_1 is the fixed effect coefficient for the Condition variable. - Condition$_{ij}$ is the value of the Condition variable for the i-th observation for the j-th participant. - u_{0i} is the random intercept for the i-th participant, drawn from a normal distribution with mean zero and some participant-specific variance. - u_{1j} is the random intercept for the j-th Hazard Clip, drawn from a normal distribution with mean zero and some Hazard Clip-specific variance. - e_{ij} represents the residual error term for the i-th observation for the j-th participant. The fixed effects are denoted by β coefficients, and the random effects are represented by u terms.

Tablet HDD *(Est. = -1.64, 95% CI [-2.15, -1.12], p < .001)* conditions were significantly lower than Baseline confidence ratings. After refactoring the model to use Cued AR HUD as the intercept, confidence ratings in the AR HUD *(Est. = -0.53, 95% CI [-1.01, -0.06], p = .027)*, AR HDD *(Est. = -1.27, 95% CI [-1.74, -0.80] p < .001)*, and Tablet HDD *(Est. = -1.04,95% CI [-1.51, -0.57] p < .001)* conditions were significantly lower than the Cued AR HUD condition. Refactoring with AR HUD as the intercept, confidence ratings in the AR HDD *(Est. = -1.27, 95% CI [-1.74, -0.80], p < .001)* and the Tablet HDD (*(Est. = -1.04, 95% CI [-1.51, -0.57], p < .001)* were significantly lower than the AR HUD condition. There were no significant differences between confidence ratings for the AR HDD and the Tablet HDD conditions (see Table 5.5).

Condition	Confidence Rating	Std Err	Lower CI	Upper CI
Baseline	0.86	0.029	0.799	0.914
AR HUD	0.68	0.045	0.583	0.76
Cued AR HUD	0.78	0.038	0.698	0.847
AR HDD	0.5	0.05	0.405	0.599
Tablet HDD	0.565	0.05	0.46	0.652

Table 5.4: Summary statistics for the average confidence rating for each Presentation method in Experiment 3, as well as the standard error and both lower and upper confidence intervals as reported from the mixed effects model.

	Comparison														
Confidence	Baseline			AR HUD			Cued AR HUD			AR HDD			Tablet HDD		
Model Intercept	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Baseline	1.87	(SE = 0.25)		-1.12	(SE = 0.27)	p <.001***	-0.59	(SE = 0.28)	p = .031*	-1.86	(SE = 0.26)	p <.001***	-1.63	(SE = 0.26)	p <.001***
HUD				0.74	(SE = 0.21)		0.53	(SE = 0.24)	p = .027*	-0.74	(SE = 0.22)	p = .001**	-0.51	(SE = 0.22)	p = .023*
Cued AR HUD							1.28	(SE = 0.22)		-1.27	(SE = 0.24)	p <.001***	-1.04	(SE = 0.24)	p <.001***
HDD										0.01	(SE = 0.2)		0.23	(SE = 0.24)	p = .3
Tablet HDD													0.24	(SE = 0.2)	

Table 5.5: Model estimates for Confidence ratings, with the Standard Error (SE) and p values obtained through Wald's approximation for each of the GLME models for each of the 5 Presentation Methods. Each row corresponds to a model with the named presentation condition as the intercept, the column representing each of the other presentation conditions compared to the intercept. Repeat comparisons are omitted for clarity, but represent the inverse of the estimate presented.

Subjective Attention Ratings

Since the Subjective Attention ratings were only recorded at the end of each condition, the data were not nested and thus a mixed effects model was not suitable. A mixed design ANOVA with participant as a random effect found a significant difference between the Presentation Methods $(F\ (4,\ 92) = 21.5,\ p < .001,\ \eta^2 = 0.35)$. *Post hoc* analyses with a Bonferroni adjustment revealed that Attention ratings in all presentation conditions were significantly lower *(p < .001)* than Baseline in all conditions. However, none of the comparisons between conditions were significantly different.

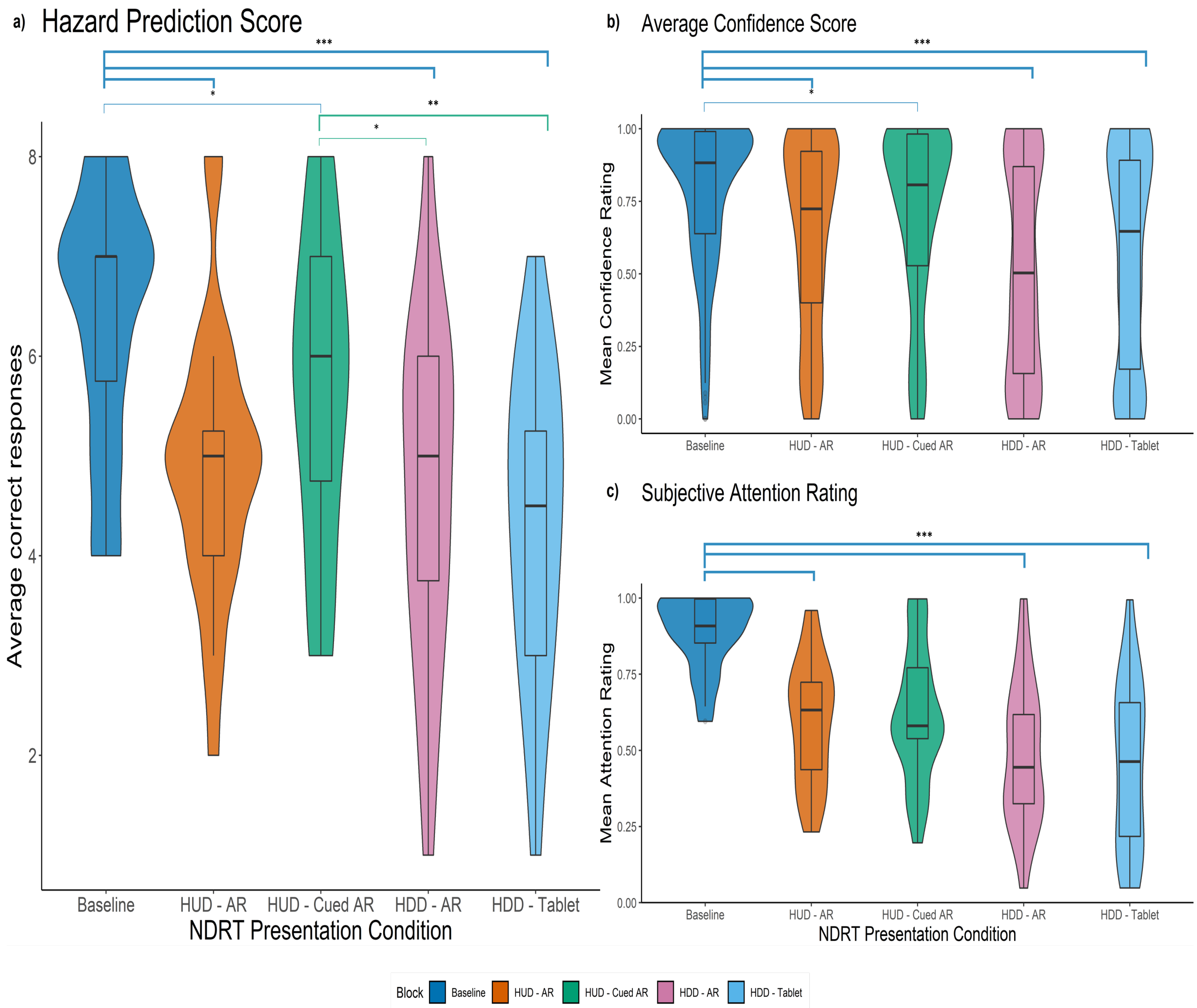


Figure 5.4: Average scores on the Hazard Prediction task (left) average confidence ratings (top) and subjective attention ratings (bottom) in each Presentation Method condition for Experiment 3. Each of the conditions showed a significant decrease in both average score, confidence and subjective attention ratings in all of the NDRT presentation conditions compared to Baseline.

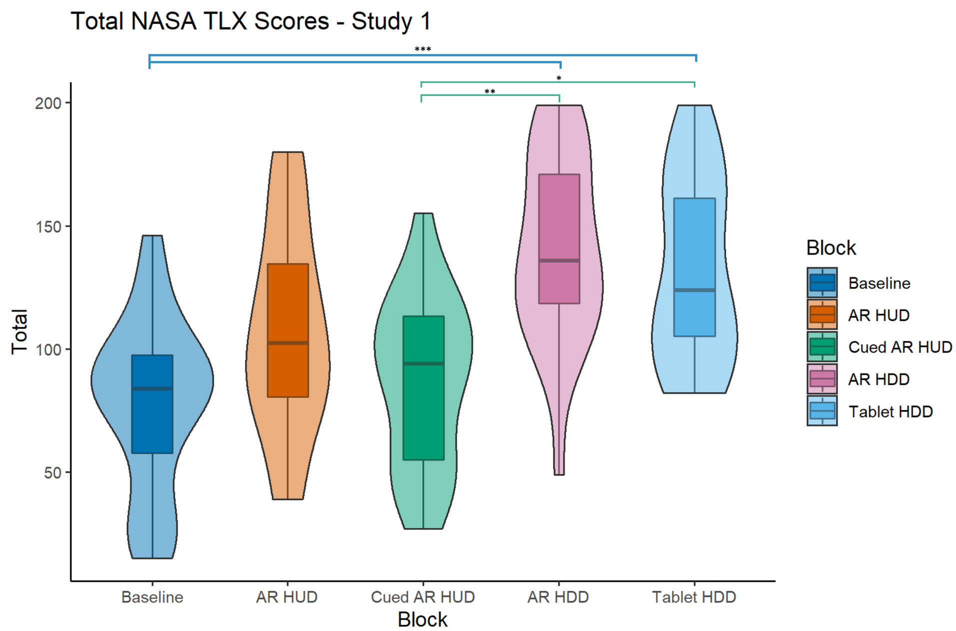


Figure 5.5: Raw Total NASA TLX scores for each NDRT presentation condition for Experiment 3.

Perceived Workload

A mixed design ANOVA was conducted on the raw total NASA TLX ratings at each condition. The Total TLX Score was statistically significantly different across different conditions *($F(2.91, 64.02) = 20.59$, $p < .001$, $\eta^2 = 0.31$)*. *Post hoc* analyses with a Bonferroni adjustment revealed that the pairwise comparisons between the Baseline condition ratings and both the AR HDD *($p < .001$)* and the Tablet HDD *($p = .004$)* conditions were significantly different, but not the AR HUD or Cued AR HUD conditions (see Figure 5.5). There were also significant differences between the Cued AR HUD condition and both the AR HDD *($p = .002$)* and the Tablet HDD *($p = .03$)*. No other comparisons between conditions were significantly different however. For full breakdown of NASA TLX scores, subscales and comparisons (see Appendix C).

Simulator Sickness Scores

Sickness ratings were compared between the 5 timepoints and the baseline rating. A LME model to predict SSQ ratings across each timepoint. The model included participant as random effect and its total explanatory power was substantial (conditional $R^2 = 0.7$) with the formula:

$$\text{Total SSQ Score}_{ij} = \beta_0 + \beta_1 \cdot \text{Timepoint}_i + u_j + \epsilon_{ij}{}^3$$

This model was compared to a null model with Timepoint and participant fitted as random effects, and was found to explain a significantly greater proportion of the variation *(Null Model AIC = 1053.7, BIC = 1062.6; SSQ Model AIC = 1044.8, BIC = 1068.6; p = .002)*. Within this model, Total SSQ scores at Timepoint 4 were significantly higher than those at Baseline *(Est. = -4.99, 95% CI [-8.93, -1.04], t(136) = -2.50, p = 0.014)*, Timepoint 1 *(Est. = -5.30, 95% CI [-9.24, -1.35], t(136) = -2.66, p = 0.009)* and Timepoint 2 *(Est = -4.52, 95% CI [-8.46, -0.57], t(136) = -2.27, p = 0.025)*. Furthermore, SSQ Scores at Timepoint 5 were significantly higher than those at Baseline *(Est. = -6.52, 95% CI [-10.47, -2.56], t(136) = -3.26, p = .001)*, Timepoint 1 *(Est. = -6.83, 95% CI [-10.79, -2.87], t(136) = -3.41, p < .001)*, Timepoint 2 *(Est. = -6.05, 95% CI [-10.01, -2.09], t(136) = -3.02, p = .003)*, and Timepoint 3 *(Est.= -4.8, 95% CI [-8.76, -0.84], t(136) = -2.40, p = .018)*. No other model comparisons produced significant difference (see Table 5.6 for an abbreviated table, and Table C.3 for the full list of Model comparison for Total SSQ score).

Total SSQ Scores	T4			T5		
Model Intercept	Est	SE	Sig	Est	SE	Sig
Baseline	4.987	1.995	**p = .014***	6.516	2.001	**p = .001****
T1	5.298	1.995	**p = .009****	6.828	2.001	**p < .001*****
T2	4.519	1.995	**p = .025***	6.049	2.001	**p = .003****
T3	3.273	1.995	p = .105	4.802	2.001	**p = .018***
T4	12.311	2.517	-	1.53	2.001	p = .45

Table 5.6: An abbreviated table showing the Model Estimates, Standard Error (SE) and p values obtained through Wald's approximation for each of the LME models of Total SSQ scores at each timepoint of the experiment. Only the Model intercepts with significant differences are included, and the main table can be found in Table C.3.

[3] β_0 represents the intercept. β_1 is the coefficient for *Timepoint* . u_j is the random effect associated with *participant* j, assumed to be normally distributed with mean 0 and variance σ_u^2. ϵ_{ij} is the residual error term, assumed to be normally distributed with mean 0 and variance σ_ϵ^2.

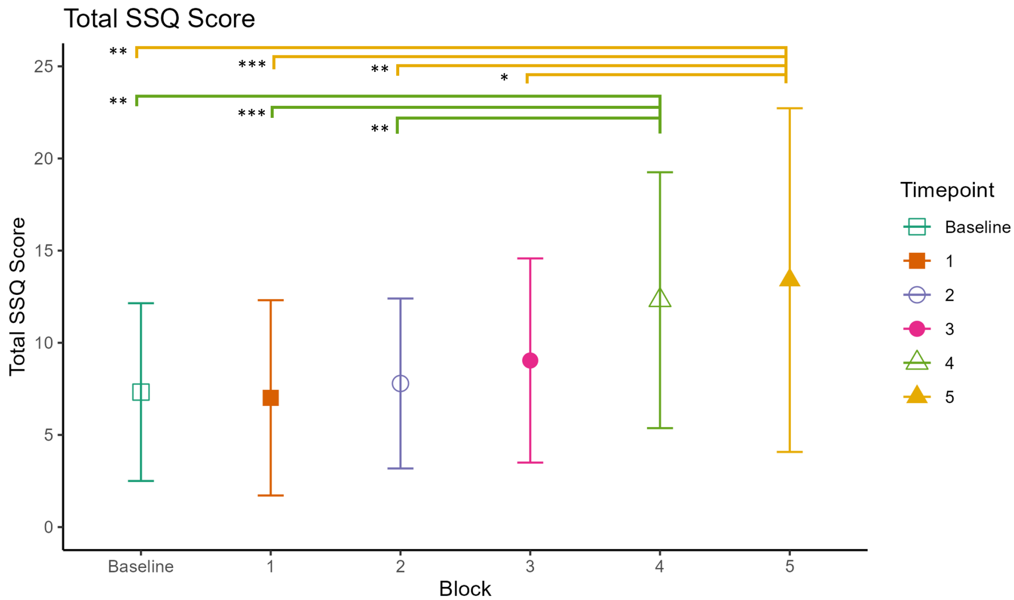


Figure 5.6: Average SSQ rating at each of the 6 timepoints of the experiment. There was a significantly higher ratings at T4 and T5 of the experiment, which is concordant with previous findings that sickness symptoms present after 20 minutes in AR (Hughes et al., 2020).

Ratings on each of the three subscales of the SSQ (Nausea, Oculomotor discomfort and Disorientation) were calculated individually and compared across the 5 conditions. LME models predicting change in these subscales with timepoint as a fixed effect and participant as random effect were fitted using the same formula as above. These models were compared to ones which included Condition as a main and an interaction effect, but these were not found to explain a significantly greater portion of the variance and so were discarded in favour of the simpler models. The model comparing SSQ scores on the Nausea subscale ratings was found not to explain the variance in scores significantly better than a null model showing no differences.

However, the model comparing Oculomotor subscale ratings (with substantial explanatory power, conditional $R^2 = 0.69$) found that ratings were significantly higher at Timepoint 4 compared to Baseline *(Est. = -6.32, 95% CI [-10.88, -1.75], t(136) = -2.74, p = .007)*, Timepoint 1 *(Est. = -6.63, 95% CI [-11.20, -2.07], t(136) = -2.87, p = .005)* and Timepoint 2 *(Est. = -5.68, 95 CI [-10.25, -1.12], t(136) = -2.46, p = .015)*. Furthermore, Oculomotor discomfort ratings were also significantly higher at Timepoint 5 compared to Baseline *(Est. = -8.17, 95% CI [-12.75, -3.59], t(136) = -3.53, p < .001)*, Timepoint 1 *(Est. = -8.48, 95% CI [-13.06, -3.90], t(136) = -3.66, p < .001)* Timepoint 2 *(Est. = -7.54, 95% CI [-12.11, -2.96], t(136) = -3.25, p = .001)* and Timepoint 3 *(Est. = -5.64, 95% CI [-10.22, -1.06], t(136) = -2.44, p = .016)*.

Additionally, a model comparing Disorientation subscale ratings (with substantial explanatory power, conditional $R^2 = 0.62$) at each timepoint found significant differences between Timepoint 4 and ratings at Baseline *(Est. = -6.38, 95% CI [-11.10, -1.66], t(136) = -2.67, p = .008)*, Timepoint 1 *(Est. = -5.80, 95% CI [-10.52, -1.08], t(136) = -2.43, p = .016)*, and Timepoint 2 *(Est. = -5.22, 95% CI [-9.94, -0.50], t(136) = -2.19, p = .03)*. Similarly, significant differences were also found between Timepoint 5 and ratings at Baseline *(Est. = -6.29, 95% CI [-11.02, -1.55], t(136) = -2.63, p = .010)*, Timepoint 1 *(Est. = -5.71, 95% CI [-10.44, -0.97], t(136) = -2.38, p = .019)*, and Timepoint 2 *(Est. = -5.13, 95% CI [-9.86, -0.39], t(136) = -2.14, p = .034)*.

Change in Simulator Sickness Scores

Sickness ratings were then converted to show change over the 6 time points of the experiment i.e., an increase in reported sickness would give a positive value, whilst a decrease would give a negative one. After converting the scores, a LME model to predict perceived task sickness score across Block was fitted comparing ratings at each the 5 timepoints to the previous timepoint. However, there were no significant differences in the change over time for subsequent ratings.

NDRT Performance

Performance on the AR NDRTs was compared [4]. This was calculated by taking the number of gems spawned during the block and counting the overall number of gems popped to get the proportion of gems that participants hit. The average scores for each condition were 47.5 % *(SD = 16.5)* for the AR HUD condition, 50.2% *(SD = 16.53)* for the Cued AR HUD condition and 22.2% *(SD = 11.2)* for the AR HDD condition. A one-way repeated measures ANOVA was conducted on the task performance measures for only the AR NDRTs. The proportion of gems popped was statistically significantly different between conditions (*F(1.56) = 80.91, p < .001,* η^2 *= 0.42). Post hoc* analyses with a Bonferroni adjustment revealed a significantly higher proportion of gems popped in both the AR HUD condition *(p < .001)* and the Cued AR HUD condition *(p < .001)* compared to the AR HDD condition.

[4] Performance on the Tablet HDD was not compared as the task had different requirements and scoring measure to the AR NDRTs.

Participant Rankings

In post experiment interviews, participants were asked to rank their preferred condition for completing the driving task in order of most to least favourite. These were converted into positional rankings for each user, where most preferred equalled to 5 points and least preferred equalled to 1 point. The rankings were then summed and weighted by the number of points each ranking received. The Baseline condition with no non-driving task was ranked the highest, followed by the Cued AR HUD condition and then the AR HUD condition. Both the HDD conditions were ranked the lowest, with the AR HDD task being the least favoured. See Table 5.7 for the full list of rankings.

Condition	1st	2nd	3rd	4th	5th	Ranking
Baseline	13	6	3	2	0	**102**
AR HUD	2	6	13	2	0	77
Cued AR HUD	6	13	5	2	0	99
Tablet HDD	2	1	2	12	7	51
AR HDD	0	0	0	7	17	31

Table 5.7: Participant rankings of each Presentation Method, with the number of points gained for each ranking. Weighted rankings were calculated by multiplying the count by the number of points next to each rank (5 for 1st, 4 for 2nd etc) to show that the Baseline condition was the most preferred by participants.

5.5 Experiment 3 Discussion

Results from this experiment showed that participants were able to maintain situational awareness of the driving task whilst engaged with an NDRT. Hazard prediction scores in all conditions were higher than chance, indicating the participants were able to use their driving experience to correctly predict what happened next in the hazard clips. However, there was no clear benefit observed when presenting the NDRT via an AR HUD compared to the HDD conditions, contrary to what previous research into AR HUDs (Lindemann et al., 2018; Jing et al., 2022) may suggest. This indicates that the distracting nature of a NDRT hinders drivers' abilities to monitor the road, regardless of the presentation condition. Only when an attentional cue indicating the location of the hazard was included in the AR HUD was there any benefit to situational awareness. However, this was still lower than when participants focused solely on the driving task in the Baseline condition. Therefore, the presentation of NDRTs to drivers that are still required to maintain supervision of an AV does not benefit from a HUD presentation.

Real World Comparison - Baseline vs Tablet HDD

The Tablet HDD condition was included to offer a real-world comparison with Bejeweled, a commercial application with 25 million users (Smith, 2023), used as the NDRT. This type of task is currently legal in the UK in L3 vehicle in autonomous mode, so this comparison represented the effects of using a real application likely to be used in an AV. The hazard prediction scores in this condition were significantly lower than the Baseline condition. This is significant since car manufacturers currently employ HDDs for their in-car infotainment systems (Gold, 2022). Despite UK law stating that these types of displays are legal, results from this experiment suggest that this type of NDRT presentation is detrimental to situational awareness, and contradicts the *"do not prevent the driver from responding to demands from the automated driving system"* requirement from the UK & Scottish Law (UK and Scottish Law Commissions, 2020). Though specific comparisons cannot be made from this experiment, the results suggest presenting the NDRT using an AR HUD with an attentional cue towards dangers on the road is better for driver awareness than current HDD methods.

Effect of the Attentional Cue

Including an attentional cue in the Cued AR HUD did lead to better prediction scores compared to the HDD conditions. However, there was no difference when compared to the HUD without a cue, and it was worse than full attention to the road. These results may be because the design of the attentional cue not effectively communicating a warning to participants.

The red gem used as a cue here did not disrupt performance in the NDRT and only appeared on the screen near the area of the hazard, with only its presence signalling a hazard. Compared to other attentional alerts that are multimodal (Politis et al., 2014b, 2017), the cue was not as attention capturing. This may explain why there were no significant differences between the Cued AR HUD and AR HUD conditions; the AR content in both of these conditions did not obstruct view of the road, similar to the Pokémon Drive concept from Schroeter and Steinberger (2016), or the zombie shooting game from Togwell et al. (2022) . However, this unintrusive design does not necessarily reflect the type of tasks that people report wanting to engage with as NDRTs. Productivity tasks such as answering emails, browsing the internet or watching films are all popular suggestions as NDRTs that AVs will allow (Panagiotopoulos and Dimitrakopoulos, 2018; McGill et al., 2017) which would be more obstructive of the road view. The results from this experiment show how a HUD presentation is not necessarily beneficial for situational awareness as posed by RQ2, but does not further understanding of the complexities behind how including attentional cues in NDRTs affects driver awareness, put forward in RQ3.

Task Design Changes and Expert Opinions on Cue Design

Results from this experiment suggest a visual attentional cue is needed to be incorporated into the NDRT for a HUD presentation to aid situational awareness. Yet from the design of this experiment there are still outstanding questions:

- Do these results persist with a more realistic NDRT?
- Does the design of the attentional cue impact on how it aids situational awareness?

To further explore these questions, a follow-up experiment was designed to expand on the results collected here. A keypad dialling task was selected as to represent the type of text input tasks drivers engage with in cars, as well as being a task used in similar research evaluating input devices for NDRTs (Large et al., 2016; Jung et al., 2021). Furthermore, to make more specific comparisons between HUDs and HDDs, a task that required the same manner of interaction was needed to evaluate how the specific demands of the NDRT impacted Hazard Prediction ability. The HUD conditions in Experiment 3 only required eye-gaze to perform, whereas the HDD conditions required a touch input. These different input methods may explain the significant difference in results beyond the change in Presentation Method.

It was deemed necessary to include a more dynamic attentional cue which drew attention to not just the road, but a specific location of interest such as a dangerous hazard. Given the novelty of engaging with an NDRT whilst supervising an AV, there are few examples of how such a cue could be designed to impart situational awareness that does not involve full attention capture. A group of Automotive UI and HCI experts were consulted to gain insight into possible cue designs. This was with the focus on making use of the dynamic aspects of an AR HUD to communicate information to the driver. Design sessions to create informative attentional cues raised the importance of colour change and motion to indicate the location of hazards, as well as the levels of danger associated with them. These designs ideas were consolidated and implemented into a keypad dialling task to create a Dynamic AR HUD condition, signalling to the participants areas of the road where a hazard was located.

5.6 Expert Focus Group for Designing Dynamic Cues

Given the novelty of engaging with an NDRT whilst supervising an AV, there are few examples of how an attentional cue could be designed to impart situational awareness information that does not involve full attention capture. As such, a focus group was conducted with experts in the design of Mixed Reality and in-car displays to gather a wider informed opinion on how these attentional cues could be designed.

Participants

Six expert participants *(2 Female, 4 Male)* were recruited for the focus group. They were recruited from the University of Glasgow School of Computing Science, and were comprised of three Automotive UI and three HCI researchers (determined by having published in these respective fields). Three held a driving license from the UK, 1 from Israel, 1 from Germany and 1 was learning to drive. The average age was 35.4 years *(min 24, max 56)* and the average driving experience was 15.6 years *(min 3 months, max 39 years)*.

Design

The focus group design was similar to that used by Al-Taie et al. (2023). Participants were given pens, chapter and laminate sheets to create their designs. The focus group lasted 45 minutes, and was structured to give participants time to develop their own designs as well as receive feedback and discuss ideas with other participants.

Procedure

Participants were provided with an information sheet and a consent form recording their demographic data. Then, participants were given a brief introduction presentation regarding AR HUDs and driver awareness in AVs. They were given 10 minutes to design a HUD which provided dynamic cues during an NDRT individually. Following this, they swapped designs with a partner to give feedback on their design for 2 minutes. To encourage evaluation of the designs, they were given 'like' and 'bomb' stickers as in Al-Taie et al. (2023) to signal which aspects of each other's designs they particularly liked or disliked. Participants then had 5 minutes to finalise their designs and address the feedback they were given, before presenting it to the whole group for discussion (see Figure 5.7 for examples of the designs). Finally, an unstructured group interview was conducted to further explore any ideas that had come up during the group discussion.

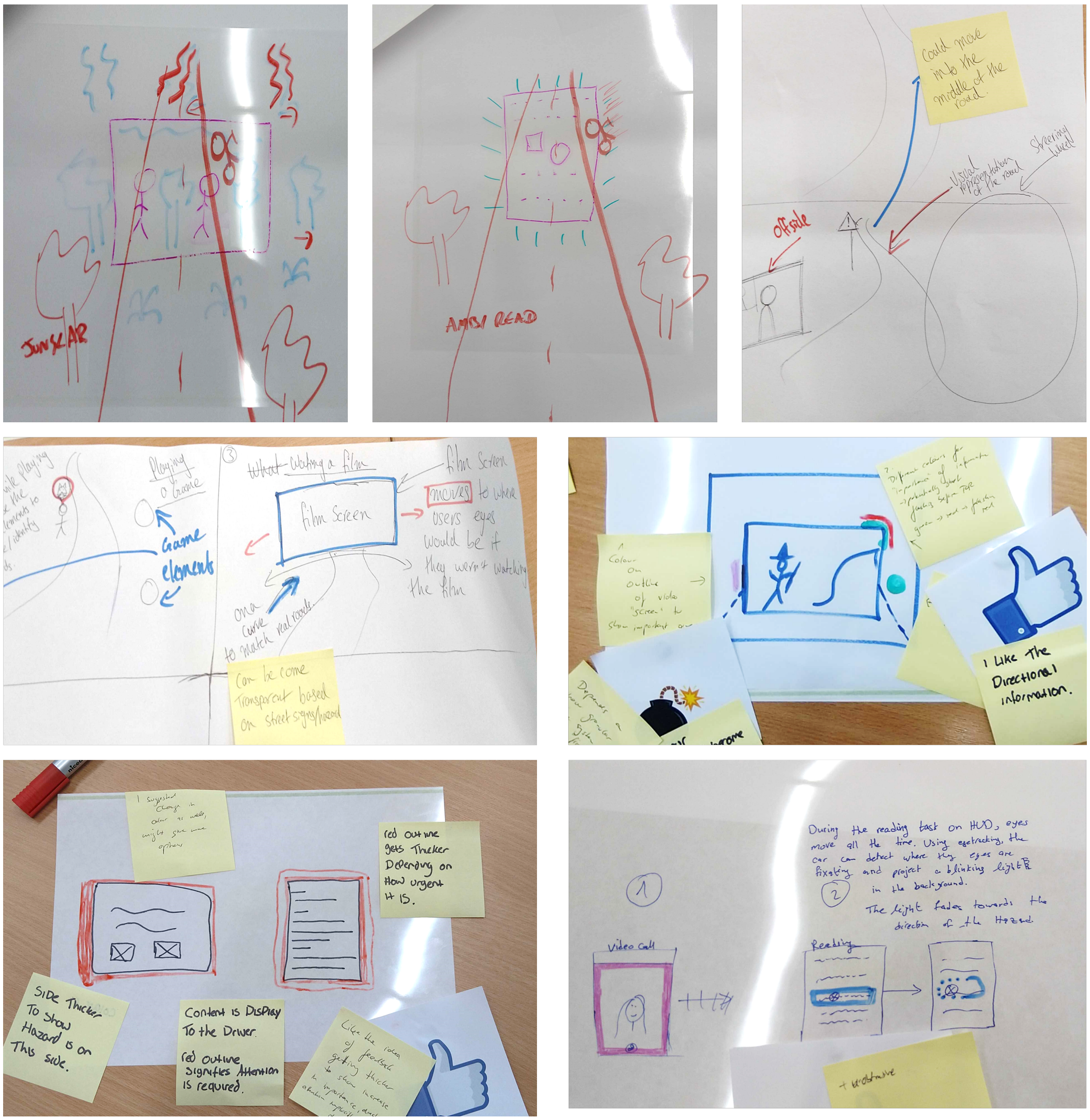


Figure 5.7: A snapshot of some of the focus group designs for dynamic attentional cues in a non-driving task.

Focus Group Results

The recording of the design explanations and subsequent discussions was transcribed and analysed. This analysis was conducted to discover the prominent ideas, concepts and designs for the attentional alerts, loosely based on the thematic analysis structure following guidelines set out by Braun and Clarke (2019) (see Table 5.8 for the full list of codes).

Theme	Codes	Discussing ideas	Unstructured interview	Total
Displaying AR content to the driver	Changing colour to signify meaning	5	4	9
	Highlight different severity of hazards	5	1	6
	Movement of AR objects	2	1	3
	Using location as an indicator	4	2	6
	Change in opacity/intensity	2	1	3
Impact on driver behaviour	Not interrupting driver behaviour	3	3	6
	Explicitly attracting attention	3	0	3
	Measuring driver awareness\state	1	0	1
	Not interrupting NDRT	0	1	1
	Confusion between AR and reality	0	2	2
How to impart information via cues	Bringing AR elements into the real world	1	0	1
	Other modalities are more explicit	0	4	4
	Bringing real world elements into AR	5	0	5
	Including audio as a separate cue	1	1	2
	Trust in the vehicle	0	2	2

Table 5.8: The three themes that emerged during the focus group discussion, as well as the underlying codes from the data.

From the focus group discussions, 3 themes were identified:

- Displaying AR content to the driver
- Impact on driver behaviour
- Representation of road information

Theme 1 : Displaying AR content to the driver

The most common element discussed was how AR content was displayed to the driver, with changing colour being the most common way described to convey information. Five of the six designs included an aspect of colour change to signify that a hazard was present and that the driver should pay attention to the road.

> *"You might have some kind of colours around the edge representing that (enhancing the experience), except when you come across to the bike, when that might be enhanced with red." - PD06*

> *"The actual element changes colour of like red, as, 'this is the human'. This is where the biggest hazard is, and then yellow or amber to something that could happen" - PD01*

This was associated with different colours representing different level of danger or severity associated with the hazard. In the unstructured interview when asked why certain colours were chosen, participants highlighted that conventions from existing road architecture were used in their designs to communicate similar messages and intent.

> *"I was thinking about like having different colours also like green one or something, like it's good to know but don't worry about it too much then it becomes red if it's really important and it starts flashing" - PD05*

It was noted that these heuristics are relatively, but not totally, universal and so including them as part of the HUD design makes use of a driver's existing understanding of the road.

> *"That's a convention. I think it's an International Convention. Everyone will understand that red is hazard. Green is OK." - PD04*

In addition to colour, location of the cue was also highlighted as important in communicating information to the driver. The idea of drawing the driver's attention towards specific areas of the road through highlighting areas appeared in 3 designs.

> *"So it's basically watching a movie that is not on the whole windscreen, but just a screen... and then highlighting there is an area of interest like on the side where something is happening, basically to direct your eyes there." - PD05*

> *"So using an eye tracking technique to detect where the [driver's] eyes are. So when there's a hazard somewhere ... where the (driver's) fixation is, the car or the HUD should issue, some light or some frame, ... the fading of it will be gradual and towards the direction of the hazard." - PD04*

Theme 2 : Impact on driver behaviour

Another important aspect of HUDs discussed was that it did not interrupt the driver's natural behaviour. The idea of placing content in the HUD where the driver would already be looking was raised in multiple designs in order to reduce the workload of swapping between the NDRT and the road.

> *"A head up display, makes more sense. You're already looking at something." - PD05*

This was with the aim of bringing elements of the real world into the AR display and incorporating it into the NDRT, to make attending to it as easy as possible for the driver.

> *"...it brings you to the hazard and not the hazard to you. Yeah, because obviously you know a child here (in the road scene) is there in the real world and your eyes are being brought to that." - PD01*

Conversely, there was discussion of when and how a driver's attention needs to be explicitly captured, such as in a critical takeover situation. This was related to the ideas presented above differentiating between different levels of danger with each hazard.

> *"You can look at right now to differentiate like urgency because we do this with like a few designs, do this with colour like, you know, green, amber, red. But I feel like if it's like 'a thing, look at it right now', the flashing is the way to do it." - PD01*

> *"But if there was a bicycle in the way ... you might enhance the look of (the HUD) so they've got a bit more attention, so they weren't too subtle." - PD02*

In these critical situations where it is necessary to capture driver attention, the designs employed more disruptive methods to bring attention to certain aspects of the road and explicitly point out hazards.

Theme 3 : How to impart information via cues

The most effective way to do this was agreed to be using multimodal cues, specifically audio cues. This was deemed to be more intrusive than a visual cue and thus better at capturing attention.

> *"Was also because it should be implicit. It's not if I introduce another modality like haptic. Yeah, that is fairly explicit already. It's like something is happening because it draws your attention quite differently, yeah." - PD05*

> *"I think you want to get more multimodal ..., it's easier to kind of really reserve these more attention grabbing things." - PD05*

One design included spatial audio cues:

> *"The other idea was, if you're reading a book, about library-based spatial audio cues to say you've got pedestrian might be a child whispering from that direction..." - PD03*

However, it was pointed out how this has the potential to cause confusion. Should a driver be engaged with an NDRT and not prepared for an attentional alert, the sound needs to be recognisable, or it will blend in with other noises on the road.

> *"that would work if your cognitive load isn't high. It's like the first thing that goes like your hearing. If you're like playing in at the last level of a game or something, I don't know." - PD01*

> *"I think sounds would be more difficult to localise so like when you're driving your car, there's always sound around you. The car itself, like other things, happening, sirens." - PD06*

It was of general opinion that including audio as an attentional cue would not be as effective and should be reserved for the more critical attention alerts.

Discussion and Task Design Changes

The main outcomes from the focus group highlighted many key issues to consider when designing an AR HUD. Colour and location formed important aspects for communicating information about the road to the driver, and the intensity of those can be used to show the level of danger associated with the hazard. This is consistent with the literature highlighting colour change as a salient attention capturing cue (Hollingworth and Hwang, 2013; Wang et al., 2023), and a pre-established association of the colour red with danger (Pravossoudovitch et al., 2014; Chapanis, 1994). However, it was considered detrimental to include multimodal cues for an attentional alert, as this might be confusing to the driver. Instead, multimodal cues were considered to be more effective for critical takeover alerts due to their attention capturing nature (Politis et al., 2017, 2014b). The designs produced and the discussion from the focus group were used to inform the NDRT used in Experiment 4, in particular, the design of the cue to draw attention to the driving task. Specifically, the designs around changing the colour and location of the display to indicate danger were chosen, as these were the most common ideas discussed and the simplest to prototype and evaluate. Furthermore, to evaluate these cues, the NDRT needed to be cognitively demanding enough to draw attention away from the road, with the cue bringing attention back to the road. Continual scanning of the road as the task in Experiment 3 required was not deemed suitable, as the HUD elements only appeared infrequently and for a short length of time. These designs were implemented into a keypad text input task, which was then evaluated to test their effectiveness in aiding situational awareness.

5.7 Experiment 4

Comparing the effects of a Dynamic Cue in AR HUD vs HDD presentations of an NDRT on Hazard Prediction Ability

A follow-up experiment was conducted to compare driver situational awareness, as measured through their Hazard Prediction ability, with a more realistic NDRT presented either as a HUD or a HDD and including an attentional cue with a design informed from the expert focus group. The experimental design was approved through the College of Science and Engineering Research Ethics committee (Application number #300220198). The following section reports the methods, procedure and results from this experiment, as well as a comparison with the previous experiment.

Design

Experiment 4 followed the same repeated measures experimental design as Experiment 3, with Hazard Prediction score and subjective confidence ratings as dependent variables, and Presentation Method (Control, Static AR HUD, Dynamic AR HUD and AR HDD) as independent variables. The Baseline condition from Experiment 3 was changed into the Control condition, to ensure that the results were not due to any order effects. For the Static AR HUD condition, the NDRT was presented in front of the driving task with participants required to look through the keypad to see the road. In the Dynamic AR HUD condition, the keys would change to red and would move to create a window 4 seconds before the hazard occurred, so participants could view the part of the road where the hazard was about to occur. This was to expand on the results collected during Experiment 3 by replicating the procedure with a more obstructive NDRT and a more informative design for the attentional cue. For the AR HDD condition, the static keypad was displayed down below eye level, so participants had to take their eyes off the driving task and down to interact with the keypad.

Participants

A power analysis was conducted using the *pwr* package (Champely et al., 2016) in R Studio, which determined that for a moderate effect size of 0.81, 21 participants were required for the repeated measures design. Twenty four new participants *(Mean Age = 33.1 years, SD = 13, 11 Female)* were recruited via online forums and around the University of Glasgow Computer Science and Psychology departments. All participants had normal or corrected to normal eyesight. All were required to have held a driving license for at least 2 years but, as in Experiment 3, this was not limited to drivers from the UK (11 UK, 3 Indonesia, 2 India, 2 Greece, 1 Bulgaria, 1 China, 1 France, 1 Israel, 1 Thailand & 1 USA).

The average total driving experience was 12.14 years *(min = 1, max = 54, SD = 12.5)* and the average UK driving experience for non-UK license holders was 0.8 years *(min = 0, max = 8, SD = 2.24)*. Twelve people reported experience driving around the Glasgow area where the hazard clips were filmed, with an average experience of 5.9 years *(min = 0, max = 28, SD = 10.3)*. Five reported having used an AR headset before, ten reported using mobile AR and nine reported never having used AR, but had heard of it. Participants who took part in Experiment 3 were excluded from taking part.

Materials

Participants were presented with the same 40 video clips from the Hazard Prediction test used in Experiment 3 in blocks of 10 (see Chapter 4). Participants saw these clips in 4 Presentation Method conditions: Control, Static AR HUD, Dynamic AR HUD and AR HDD with a counterbalanced presentation order across all participants.

Augmented Reality Keypad Task

The NDRT was taken from previous work investigating distraction and mental workload in cars. A keypad dialling task similar to ones used by Large et al. (2016) & Jung et al. (2021) was taken and adapted to be displayed in the AR headset. This was developed in Unity (version 20203.26f1) using the Mixed Reality Toolkit (MRTK -version 2.7.2) and were presented using the HoloLens 2 AR headset.

For the Static AR HUD task, the keypad was displayed to participants with numbers (0-9) on top of the driving scene, partially occluding the view but with spaces and translucent textures so participants could still view the road. In the Dynamic AR HUD condition, the same AR task was presented, but the keys would move and change colour to indicate a hazard, four seconds before its onset, based on findings of average takeover request times from Eriksson and Stanton (2017). The keys would move depending on the participant's head position in relation to the position of the hazard on screen, meaning the keypad would have a slightly different layout each trial but would always move to create a consistent size window to view the hazard. The AR HDD condition displayed the same keypad from the Static condition, but down below the level of the monitor drawing participants' eyes away from the view of the road (see Figure 5.8). Participants' reaction time for dialling each 11-digit number correctly, errors, total number of correctly typed numbers, and total keypresses were measured to evaluate their performance on the task. As in Experiment 3, participants were offered a £10 incentive if they performed the best at both tasks out of all other participants to incentivise attention to both tasks.

Figure 5.8: The Keypad NDRT used in Experiment 4, showing a) the Static AR HUD, b) the Dynamic AR HUD with keys moved to show the hazard approaching and c) the AR HDD implementations

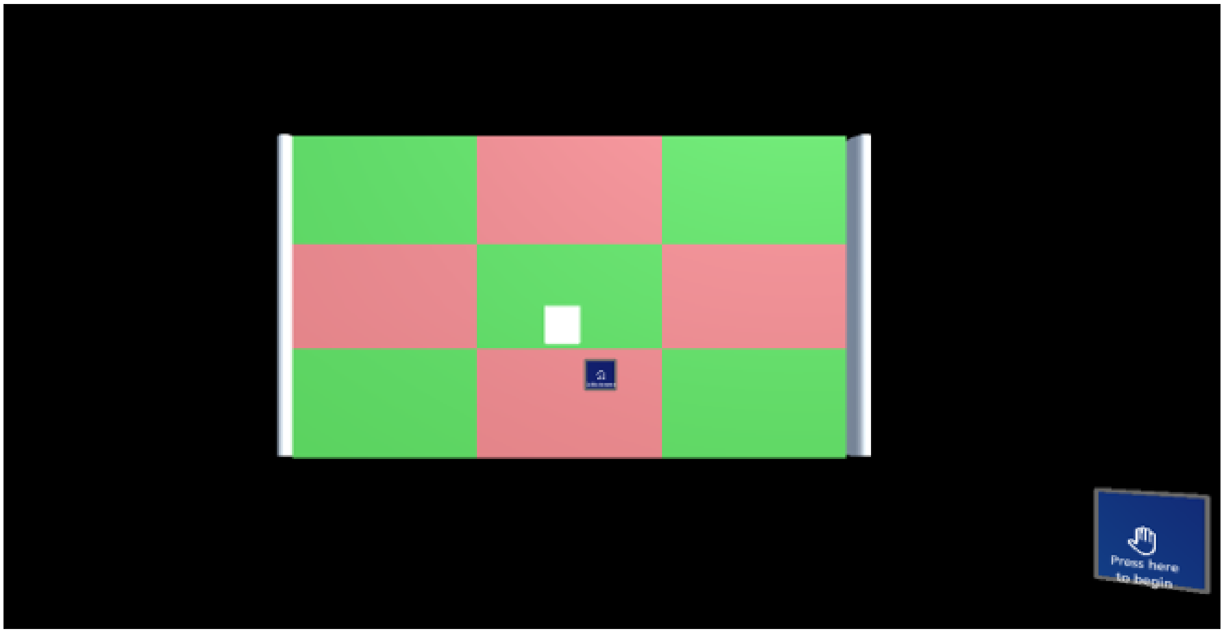


Figure 5.9: Image of the Screen AOIs in Unity 3D, which captured areas on the real-life monitor where participants were fixating. The small white cube was used by participants to calibrate the screen by aligning the two white pillars in AR with the edges of the real computer monitor.

Eye tracking

Eye movement data were collected using the Augmented Reality Eye Tracking Toolkit - (ARETT) package for Unity 3D Kapp et al. (2021) using the HoloLens 2 AR headset. A Screen object was placed in the scene made up of 9 AOI components hidden from view, which split the screen into 9 distinct areas (see Figure 5.9). Participants were asked to calibrate the location of the virtual screen in the headset by adjusting the layout of the AR scene. This was done by moving a small floating cube which acted as a parent to the Screen AOIs, and lining up two virtual pillars with the real computer monitor.

This was confirmed by the experimenter before the trials commenced (see Figure 5.9). All AR elements were labelled as AOIs using the 'Eyetracking' layer in ARETT, which logged when it was fixated on. An AOI hit was measured when the Eyetracking object collided with an object in the Unity scene, with the location of the object, its name and the duration the objects were colliding was recorded (see Kapp et al. (2021) for a detailed description of the approach used).

Procedure

The same experimental procedure as Experiment 3 was adapted for Experiment 4 (see Figure 5.3). After consenting to take part, participants provided demographic and driving experience information via the online questionnaire platform Qualtrics (Qualtrics, 2024). They were then shown an example hazard prediction clip and practised giving their response. Participants were also given a practice interaction with the headset to get used to the mid-air touch gestures, which involved pressing a button labelled from A to E that corresponded with a target letter. Participants saw the WHN clips in the 4 different presentation conditions in a counterbalanced order in blocks of 10. After watching each clip, participants were asked to predict what happens next from the list of multiple-choice answers. At the end of the experiment, they were interviewed to explain their strategies for completing the task and their preferred display for completing the experiment.

On top of the £10 voucher compensation for taking part, participants were told they could win an extra £5 voucher if they performed the best in both the NDRT and the WHN task out of their peers to incentivise attention to both tasks. The experiment took around 60 minutes to complete and was approved by the institution's Ethics Committee.

5.8 Experiment 4 Results

The results were analysed using the same methods as Experiment 3. The same model fit procedure as Experiment 3 was followed (see section 5.4). The inline stats report the estimated difference and the significance for the comparison between the stated model intercept and the comparison effect (*e.g., with Condition A as intercept, scores were significantly smaller (Est. -0.66, p = .004) than Condition B*). This highlights the estimated difference between measures in the model as reported. Full model output tables show the complete set of comparisons for each dependent variable.

Hazard Prediction scores

Average scores for the Hazard Prediction task for each of the 4 conditions (Static AR HUD, Dynamic AR HUD, AR HDD and Control) were compared (see Table 5.9). A GLME model was fitted to predict the main effects of Condition on Hazard Prediction score with a random intercept for each participant and a random intercept for each Hazard Clip, with the same formula as Experiment 3 (see section 5.4).

The model's total explanatory power was moderate (conditional $R^2 = 0.38$). This model was found to explain significantly greater variance than a null model with both Condition, Hazard clip and participant fitted as random effects *(Null Model AIC = 1109.2, BIC = 1123.8; Mixed Effects Model AIC = 1091.9,BIC = 1121.1;p < .001)*. Within this model, the score for the Static AR HUD *(Est. = -0.98, 95% CI [-1.43, -0.53], p < .001)*, Dynamic AR HUD *(Est. = -0.66, 95% CI [-1.11, -0.21], p = .004)*, HDD *(Est. = -0.96, 95% CI [-1.41, -0.50], p < .001)* conditions were significantly lower than scores in the Control condition (see Figure 5.10). Refactoring the model with Static AR HUD, Dynamic AR HUD or AR HDD as intercepts produced no significant differences not already accounted for in the model described above (see Table 5.10 for a full list of model comparisons).

Condition	Score *(P)*	Std Err	Lower CI	Upper CI
Control	0.81	0.05	0.75	0.89
Static AR HUD	0.61	0.07	0.47	0.73
Dynamic AR HUD	0.68	0.06	0.55	0.79
AR HDD	0.61	0.06	0.48	0.73

Table 5.9: Summary statistics for average probability of a correct Hazard Prediction scores for each Presentation method in Experiment 4, as well as the standard error and both lower and upper confidence intervals as reported from the mixed effects model.

	Comparison											
Condition	Control			HUD			Dynamic HUD			HDD		
	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Control	1.42	(SE = 0.29)		-0.98	(SE = 0.23)	**p <.001*****	-0.66	(SE = 0.23)	p = .004	-0.96	(SE = 0.23)	**p <.001*****
HUD				0.44	(SE = 0.27)		0.32	(SE = 0.22)	p = .145	0.02	(SE = 0.22)	p = .91
Dynamic HUD							0.76	(SE = 0.28)		-0.33	(SE = 0.23)	p = .17
HDD										0.46	(SE = 0.27)	

Table 5.10: Model Estimates for Hazard Prediction Scores for Experiment 4, with the Standard Error (SE) and p values obtained through Wald's approximation for each of the GLME models for each of the 4 presentation conditions.

Confidence Ratings

Average scores for Confidence ratings were compared for each of the NDRT presentation conditions (see Table 5.12). A GLME model was fitted to predict the main effects of Condition with a random intercept for each participant and a random intercept for each Hazard Clip, with the same formula as Experiment 3 (see Table 5.4).

The model's total explanatory power was moderate *(conditional $R^2 = 0.34$)* and was found to explain significantly greater variance than a null model with Condition and participant fitted as random effects *(Null Model AIC = 1118.8, BIC = 1133.4; Mixed Effects Model AIC = 1075.3,BIC = 1104.5;p < .001)*. Within this model, the score for the Static AR HUD *(Est. = -1.35, 9% CI [-1.83, -0.87], p < .001)*, Dynamic AR HUD *(Est. = -0.81, 95% CI [-1.29, -0.33], p < .001)*, and HDD *(Est. = -1.50, 95% CI [-1.98, -1.02], p < .001)* conditions were significantly lower than confidence ratings in the Control condition.

After refactoring the model to use Dynamic AR HUD as the intercept, Confidence ratings in the Static AR HUD *beta = -1.35, 95% CI [-1.83, -0.87], p < .001)* and AR HDD *(Est. = -1.50, 9% CI [-1.98, -1.02], p < .001)* conditions were found to be significantly lower than the Dynamic AR HUD condition. There were no significant differences between confidence ratings for the Static AR HUD and AR HDD conditions (see Table 5.11 for a summary and Table 5.12).

Condition	Rating (%)	Std Err	Lower CI	Upper CI
Control	0.85	0.04	0.75	0.91
Static AR HUD	0.59	0.07	0.46	0.71
Dynamic AR HUD	0.71	0.06	0.59	0.81
AR HDD	0.56	0.07	0.42	0.68

Table 5.11: Summary statistics for the average confidence rating for each Presentation method in Experiment 4

	Comparison											
Confidence	Control			HUD			Dynamic HUD			HDD		
	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Control	1.73	(SE = 0.3)		-1.35	(SE = 0.24)	p <.001***	-0.81	(SE = 0.25)	p <.001***	-1.5	(SE = 0.24)	p <.001***
HUD				0.38	(SE = 0.27)		0.54	(SE = 0.22)	p = .017*	-0.15	(SE = 0.21)	p = .48
Dynamic HUD							0.92	(SE = 0.28)		-0.33	(SE = 0.23)	p = .17
HDD										0.23	(SE = 0.27)	

Table 5.12: Model estimates for Confidence Ratings for Experiment 4, with the Standard Error (SE) and p values obtained through Wald's approximation for each of the GLME models for each of the 4 presentation conditions.

Subjective Attention Ratings

A mixed design ANOVA was conducted on the Attention ratings. As above a mixed effects model was deemed not suitable due to the nature of the data. The Attention rating was statistically significantly different at the different time points *($F(3, 69) = 20.28$, $p < .001$, $\eta^2 = 0.28$)*. *Post hoc* analyses with a Bonferroni adjustment revealed that attention ratings in all presentation conditions were significantly lower *(all $p < .001$)* than Control in all conditions. However, no other comparisons were significantly different.

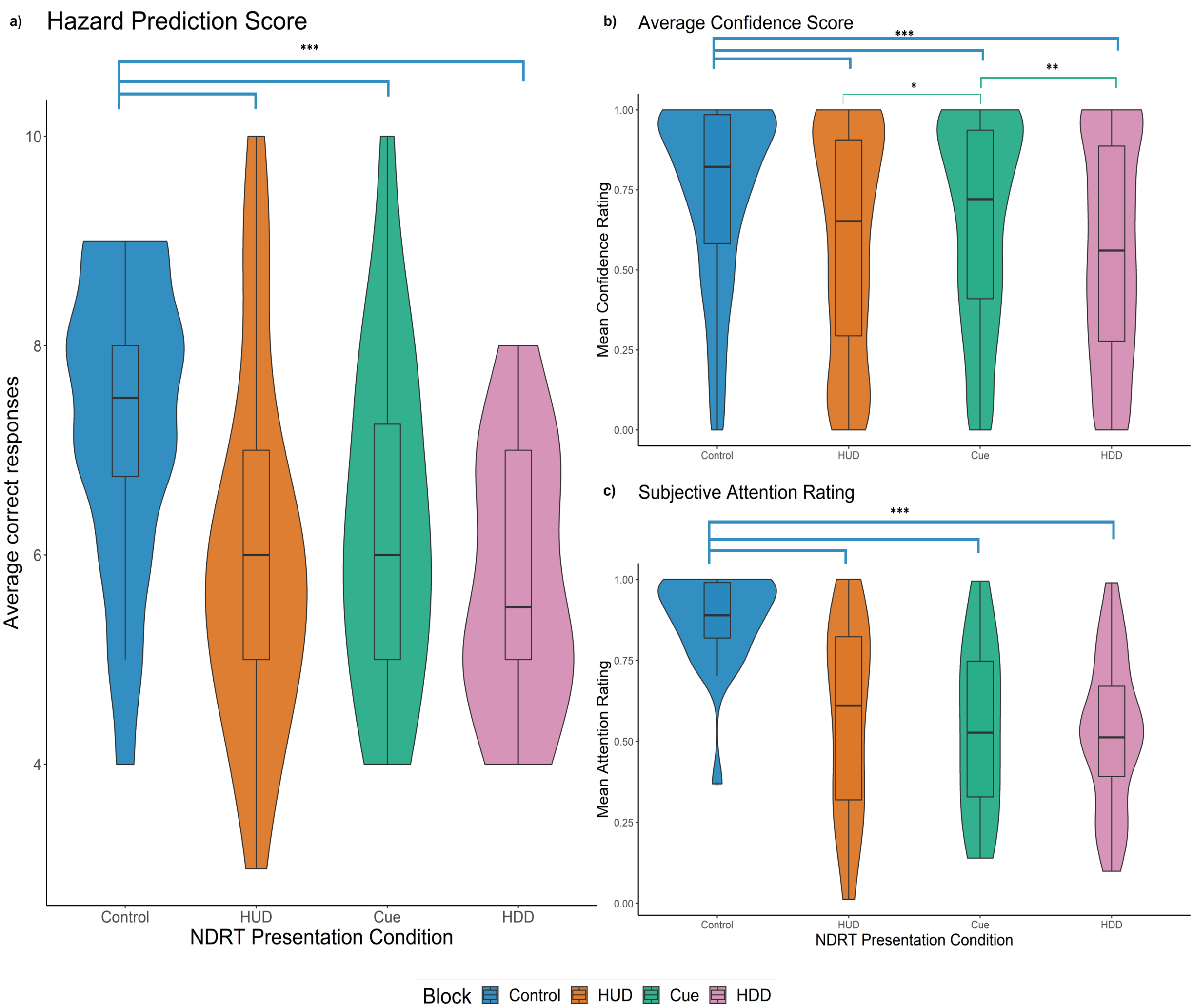


Figure 5.10: Average score on the Hazard Prediction task (left) ,average confidence ratings (top) and subjective attention ratings (bottom) for each presentation condition for Experiment 4.

Perceived Workload

A mixed design ANOVA was conducted on the total NASA TLX ratings at each condition. The Total TLX rating was statistically significantly different across different conditions, ($F(3, 69) = 39.74$, $p < .001$, $\eta^2 = 0.28$). *Post hoc* analyses with a Bonferroni adjustment revealed that the pairwise comparisons between the Control condition ratings and all other presentation conditions showed that attention ratings were significantly lower ($p < .001$) than Control in all conditions. However, none of the comparisons between the NDRT conditions were significantly different (see Figure 5.11). For each of the 6 individual TLX subscales, there were significant differences on all scales between ratings for the Control and each of the NDRT presentation conditions ($p < .001$), but no differences between each NDRT presentation condition. This was except for Effort, where the Cue condition was also rated as requiring significantly less effort than the HDD condition ($p = .023$). See Appendix C for a full list of statistical comparisons.

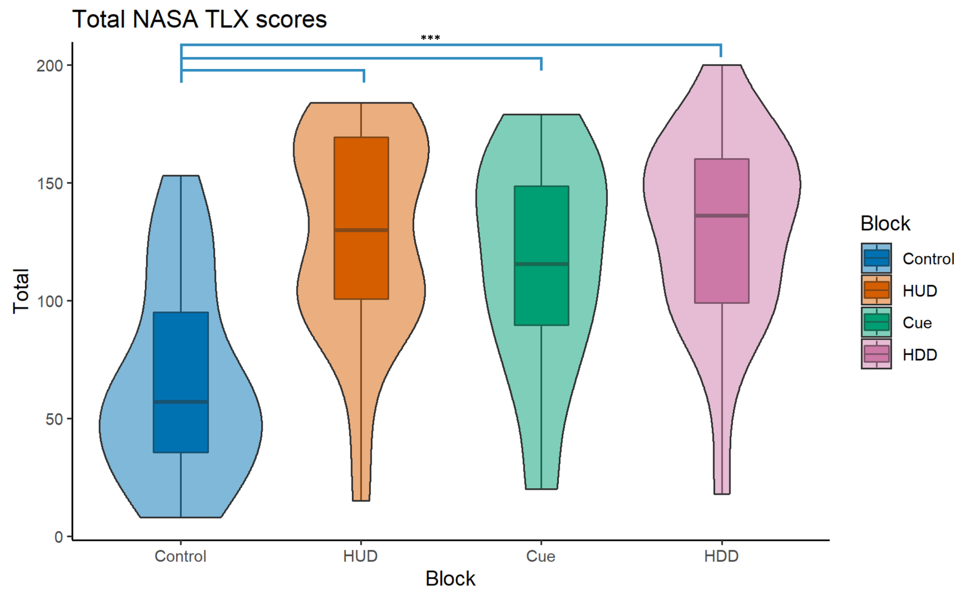


Figure 5.11: Raw Total NASA TLX scores for each NDRT presentation condition for Experiment 4.

Simulator Sickness scores

Sickness ratings were compared between the 4 timepoints and the baseline rating. A Linear Mixed Effects - (LME) model was fitted to predict SSQ ratings across each timepoint. The model included participant and condition as random effects and its total explanatory power was substantial (conditional $R^2 = 0.66$) with the same formula as Experiment 3 (see Figure 5.6). This model was found to explain significantly greater variance than a null model with Timepoint and participant fitted as random effects *(Null Model AIC = 906.95, BIC = 915.28; Mixed Effects Model AIC = 901.44 ,BIC = 920.9; p = .009).*

Within this model, Total SSQ scores at Timepoint 4 were significantly higher than those at Baseline *(Est. = -7.92, 95% CI [-12.54, -3.30], t(112) = -3.40, p < .001)* and Timepoint 1 *(Est. = -6.86, 95% CI [-11.42, -2.30], t(112) = -2.98, p = .004)* and Timepoint 2 *(Est. = -4.67, 95% CI [-9.23, -0.12], t(112) = -2.03, p = .045)*. Furthermore, Total SSQ Scores at Timepoint 3 were significantly higher than at Baseline *(Est. = -4.65, 95% CI [-9.27, -0.03], t(112) = -1.99, p = .049)*. No other model comparisons produced significant difference (see Figure 5.12, Table 5.13 for an abbreviated table, and Table D.3 for the full list of Model comparison for Total SSQ score).

Ratings on each of the three subscales of the SSQ (Nausea, Oculomotor discomfort and Disorientation) were calculated individually and compared across the 5 conditions. LME models predicting change in these subscales with timepoint as a fixed effect and participant as random effect with the same formula as Figure 5.4.

Total SSQ Scores	T3			T4		
Model Intercept	Est	SE	Sig	Est	SE	Sig
Baseline	4.648	2.33	**p = .049***	7.92	2.33	**p < .001*****
T1	3.584	2.28	p = .16	6.86	2.3	**p = .004****
T2	1.403	1.952	p = .12	4.68	2.3	**p = .045***
T3	9.973	2.749	-	3.27	2.3	p = 0.16

Table 5.13: An abbreviated table showing the Model Estimates, Standard Error (SE) and p values obtained through Wald's approximation for each of the LME models of Total SSQ scores at each timepoint Experiment 4. Only the Model intercepts with significant differences are included, and the main table can be found in Table D.3.

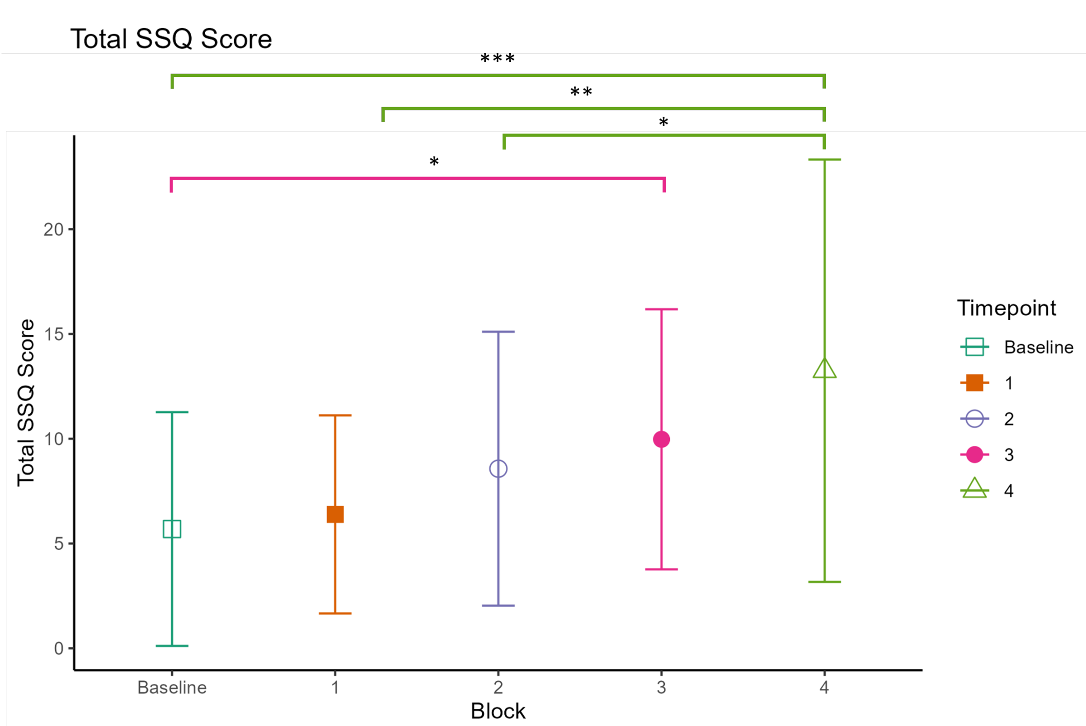


Figure 5.12: A graph showing the average SSQ rating at each of the 5 timepoint of Experiment 4. There was a significantly higher scores at T2 and T4 compared to baseline ratings.

These models were compared to ones which included Condition as a main and an interaction effect, but these were not found to explain a significantly greater portion of the variance and so were discarded in favour of the simpler models. The model comparing SSQ scores on the Nausea and Oculomotor subscale ratings were found not to explain the variance in scores significantly better than a null model showing no differences. However, the model comparing Disorientation subscale ratings (with substantial explanatory power, *conditional* R^2 *= 0.51*) found that ratings were significantly higher at Timepoint 4 compared to Baseline *(Est. = -7.69, 95% CI [-12.20, -3.19], t(112) = -3.38, p < .001)* and Timepoint 1 *(Est. = -4.64, 95% CI [-9.09, -0.19], t(112) = -2.07, p = .041)*. Ratings at Timepoint 2 *(Est. = 5.37, 95% CI [0.87, 9.88], t(112) = 2.36, p = .020)* were also significantly higher than Baseline ratings.

Change in Simulator Sickness Scores

Sickness ratings were again converted to show change over the 5 time points of this experiment. A LME model predicting perceived task sickness score across Block was fitted comparing ratings at each the 5 timepoints to the previous timepoint. However, there were no significant differences in the change over time for subsequent ratings.

NDRT Performance

A series of repeated measures ANOVAs were conducted on the task performance measures for the NDRT. The amount of numbers dialled was significantly different between conditions *($F(2,56) = 6.81$, $p < .002$, $\eta^2 = 0.07$)*. *Post hoc* analyses with a Bonferroni adjustment revealed there were significantly higher numbers dialled in the Cue condition compared to both the HUD condition *($p = .02$)* and the HDD condition *($p = .001$)*.

The number of errors was also significantly different between conditions, *($F(2,56) = 9.53$, $p < .001$, $\eta^2 = 0.16$)*. *Post hoc* analyses with a Bonferroni adjustment revealed there were significantly higher number of errors in the Cue condition compared to both the HUD condition *($p < .001$)* and the HDD condition *($p = .046$)*. However, there were no significant difference between conditions for the number of keypresses nor mean reaction times to complete a successful number dial.

Comparison with Experiment 3

To evaluate any differences between the studies, between groups ANOVAs were conducted comparing the Hazard Prediction scores, Confidence ratings, Attention ratings and NASA TLX scores between Experiment 3 and Experiment 4. NDRT Presentation Method conditions were compared with like-for-like across the two studies (Experiment 3 AR HUD - Experiment 4 Static AR HUD etc.). No significant differences were found between Hazard Prediction scores between comparable conditions, nor were there any significant differences found between confidence or attention ratings between Experiment 3 and Experiment 4. There were also no significant differences between NASA TLX scores for any of the presentation conditions, except for both of the conditions with attentional cues. A between subjects ANOVA comparing Total NASA TLX score found the workload of the Dynamic AR HUD condition of the keypad dialling task in Experiment 4 was rated as significantly higher than the Cued AR HUD condition in Experiment 3 *($F(46) = 5.85$, $p = .02$ $\eta^2 = 0.11$)*.

Eye-tracking

All analysis was conducted using functions from the ARETT R package (Kapp et al., 2021). A total of 13,335,248 gaze points were recorded across all participants in all four conditions before processing. The I-AOI fixation classification function from the ARETT R package (Kapp et al., 2021) was used to classify valid fixations. This function, based on Salvucci and Goldberg (2000)'s definition, classifies relevant fixations based on whether they land on a pre-defined AOI layer in Unity for longer than a predetermined threshold. Fixations longer than 60ms that hit either the computer screen (recorded as one AOI in a 3x3 grid - See Figure 5.9) or any 3D object in AR (a keypad button or other AR element) were retained for analysis. Given the high level of variability between participants that comes from using a portable AR headset, this method was selected over more spatially accurate classification algorithms which are sensitive to head movements, which Kapp et al. (2021) and Aziz and Komogortsev (2022) recognise as a limitation in the spatial accuracy of the HoloLens 2. After 2 participants were removed due to technical issues with the data collection, this left 538,429 valid fixations for analysis from 22 participants.

Number of Fixations

Total Fixation Count

A LME model was fitted comparing the mean Number of Fixations for each NDRT presentation condition with participants as a random effect, with substantial explanatory power *(conditional* R^2 *= 0.83)*. Overall, there were a significantly higher number of fixations in the Dynamic HUD *(Est. = 195.44, t(77) = 8.48, p < .001)* condition compared to Control and significantly fewer in the HDD condition *(Est. = -251.32, t(77) = -11.49, p < .001)*, but no differences with the Static HUD condition. There were significantly fewer fixations in both the HUD *(Est. = -167.76, t(77) = -7.36, p < .001)* and HDD *(Est. = -446.76, t(77) = -19.59, p < .001)* conditions compared to the Dynamic HUD condition. There were also significantly fewer fixations in the HDD condition compared to the Static HUD *(Est. = -279, t(77) = -12.91, p < .001)* condition (see Table 5.14 and Figure 5.13).

Hazard Fixation Count

A LME model was fitted comparing the mean number of fixations during the hazard window for each NDRT presentation condition with participant as a random factor, which had substantial explanatory power *(adj.* R^2 *= 0.74)*. Mirroring the overall fixation count results, there were a significantly higher number of fixations in the Dynamic HUD *(Est. = 16.714, t(79) = 4.83, p < .001)* condition compared to Control and significantly fewer in the HDD condition *(Est. = -33.786, t(79) = -10.29, p < .001)*.

Condition	Total Fixation			Hazard Window			Correct Fixations		
	Count	Average	SD	Count	Average	SD	Count	Average	SD
Control	6149	292.81	8.95	930	41.68	12.73	340	16.19	8.38
Dynamic AR HUD	8782	487.89	63.84	1098	44.29	12.65	332	18.4	8.57
Static AR HUD	7046	320.27	100.84	917	61	12.08	225	10.23	5
AR HDD	908	41.27	8.95	231	10.5	1.85	77	3.5	1.63

Table 5.14: Total number of fixations in each condition overall, during the Hazard Window, and the number of correct fixations during the hazard window.

There were no differences with the Static HUD condition. However, there were significantly fewer fixations in both the Static HUD *(Est. = -19.318, t(79) = -5.65, p < .001)* and HDD *(Est. = -50.5, t(79) = -14.76, p < .001)* conditions compared to the Dynamic HUD condition. Finally, there were significantly fewer fixations in the HDD condition compared to the Static HUD *(Est. = -31.182, t(79) = -9.61, p < .001)* condition (see Table 5.14 and Figure 5.13).

Correct Hazard Fixation Count

'Correct' fixations were classified depending on whether they were located in the same area of the screen as the hazard in the video clip, during the period up to four seconds before the end of the Hazard Prediction clip. This was the length of time that the buttons changed colour and location in the Dynamic HUD condition. A LME model was fitted to predict average number of fixations in Correct AOI based on Condition with participant as a random factor, which had substantial explanatory power *(conditional $R^2 = 0.56$).*

Overall, there were significantly fewer correct fixations in the Static HUD *(Est. = -5.97, t(77) = -3.39, p = .001)* and HDD *(Est. = -12.69, t(77) = -7.22, p < .001)* conditions compared to the Control condition, but no differences with the Dynamic HUD condition. There were also significantly fewer fixations in both the HUD *(Est. = -8.35, t(77) = -4.54, p < .001)* and HDD *(Est. = -15.08,t(77) = -8.19, p < .001)* conditions compared to the Dynamic HUD condition. Finally, there were significantly fewer fixations in the HDD condition *(Est. = -6.73,t(77) = -3.88, p < .001)* than in the Static HUD condition. (See Table 5.14 and Figure 5.14)

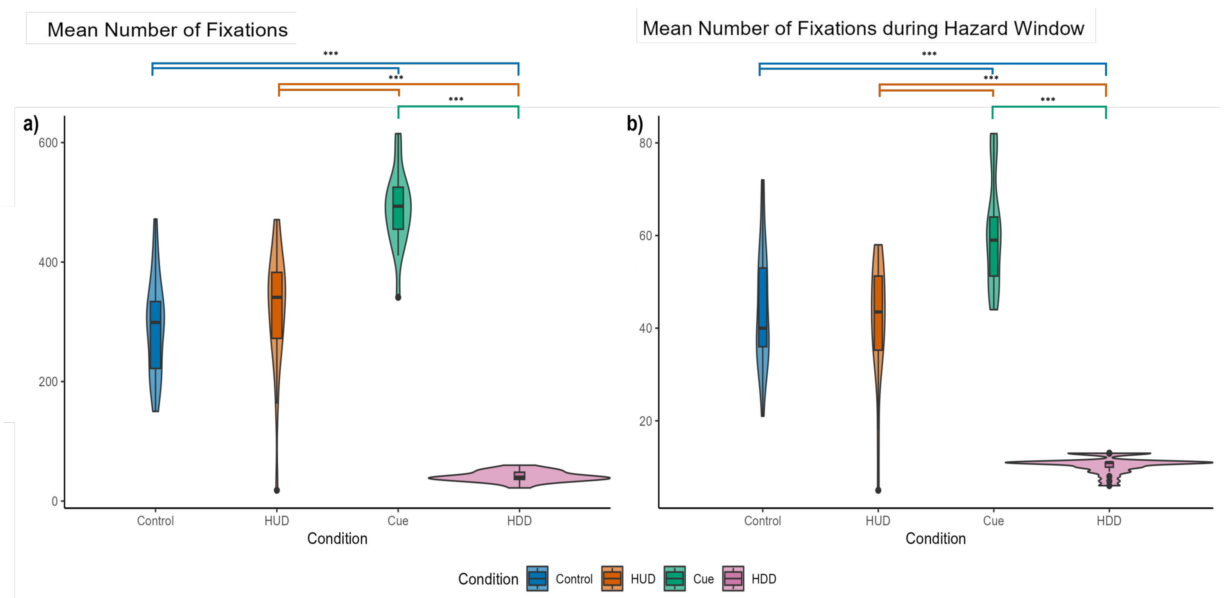


Figure 5.13: a) The mean overall number of fixations for each AR presentation condition. There were significant differences between each condition, except for between the Control and HUD conditions. b) The mean number of fixations during the Hazard Window, where the pattern of results mirrors the overall number of fixations, with differences between all presentation conditions except the Control and HUD conditions.

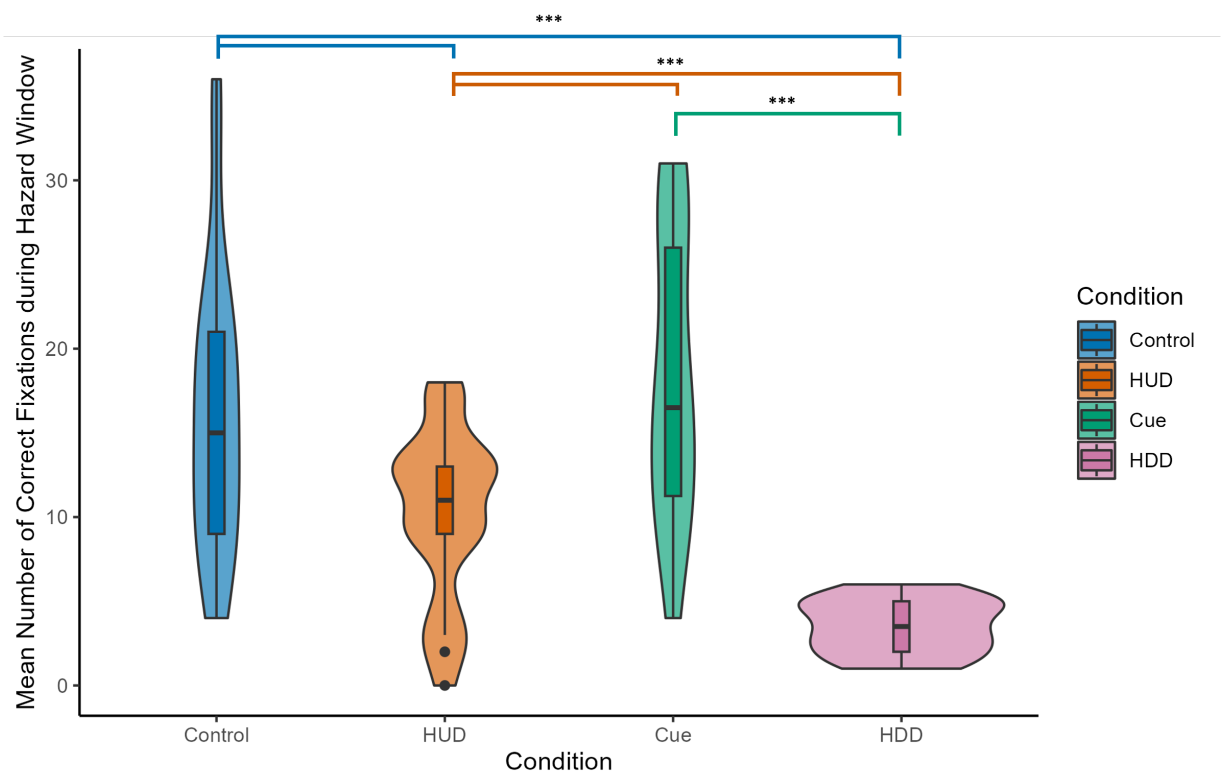


Figure 5.14: Mean number of fixations in the same AOI area where the hazard was located. There were significant differences between each presentation condition, except between the Control and Cue conditions. Both the HUD and HDD conditions had a significantly lower number of correct fixations compared to the other conditions.

Fixation Duration

Total Fixation Duration

Due to the non-normal distribution of the data, fixation duration was z-scored and values above 3 standard deviations from the mean duration were removed. Similarly, fixations which did not have a registered target or AOI target were removed, since the lack of a relevant gaze point target meant that the participant was looking at neither the screen nor any of the AR content. A LME model was fitted to model the effect of NDRT Presentation Condition on the average fixation duration with participant as a random effect that had substantial explanatory power *(conditional R^2 = 0.52)*. In the Control condition, there were significantly longer fixation durations than the Dynamic HUD *(Est. = -422.20, t(77) = -6.48, p < .001)*, Static AR HUD *(Est. = -136.53, t(77) = -2.23, p = .029)* and HDD *(Est. = -305.74, t(77) = -4.99, p < .001)* conditions. There were also significantly longer fixation durations in the Static HUD condition compared to the Dynamic HUD *(Est. =285.67, t(77) = 4.44, p < .001)* and the HDD *(Est. = -169.21, t(77) = -2.80, p = .006)* conditions (see Table 5.15 and Figure 5.15).

Condition	Mean Fixation Duration (ms)		Mean Hazard Fixation Duration (ms)	
	Mean	SD	Mean	SD
Control	1018.77	1227.32	1009.01	1138.48
Dynamic AR HUD	654.4	703.56	761.17	879.22
Static AR HUD	959.79	1024.12	1079.77	1101.62
AR HDD	836.8	959.59	790.81	807.92

Table 5.15: Mean duration of fixations in each condition overall as well as during the Hazard Window in Experiment 4.

Hazard Fixation Duration

As before, 'correct' fixations were classified depending on whether they hit the same AOI area of the screen as the hazard up to four seconds before the end of the Hazard Prediction clip. A LME model was fitted to predict Fixation length based on Condition with participants as a random effect, which had substantial explanatory power *(conditional R^2 = 0.29)*.Within this model, there was a significant difference between the fixation duration in the Dynamic HUD condition and all other conditions, with fixations being longer in the Control *(Est. = 308.65, t(77) = 2.65, p = .01)*, Static HUD *(Est. = 289.48, t(77) = 2.51, p = .014)* and HDD *(Est. = 453.42, t(77) = 3.94, p < .001)* conditions. However, there were no significant differences between any of the other conditions (see Table 5.15 and Figure 5.15).

Time to First Fixation

A LME model was fitted to model the effect of Condition on time until the first fixation on the correct area of the screen. However, there were no significant differences between any of the presentation conditions.

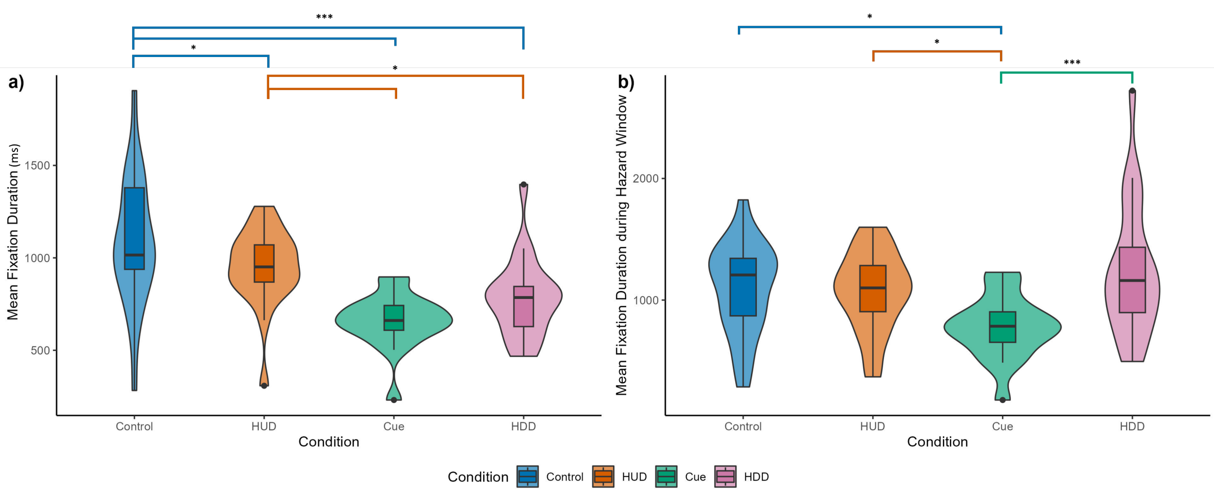


Figure 5.15: a) Mean fixation duration in milliseconds during each presentation condition. Fixations were longer in the control condition and Static HUD conditions compared to the Dynamic HUD, Static HUD and HDD conditions, but not compared to each other. b) During the hazard window, only the Dynamic HUD condition had significantly different fixations, with shorter fixations than all other conditions.

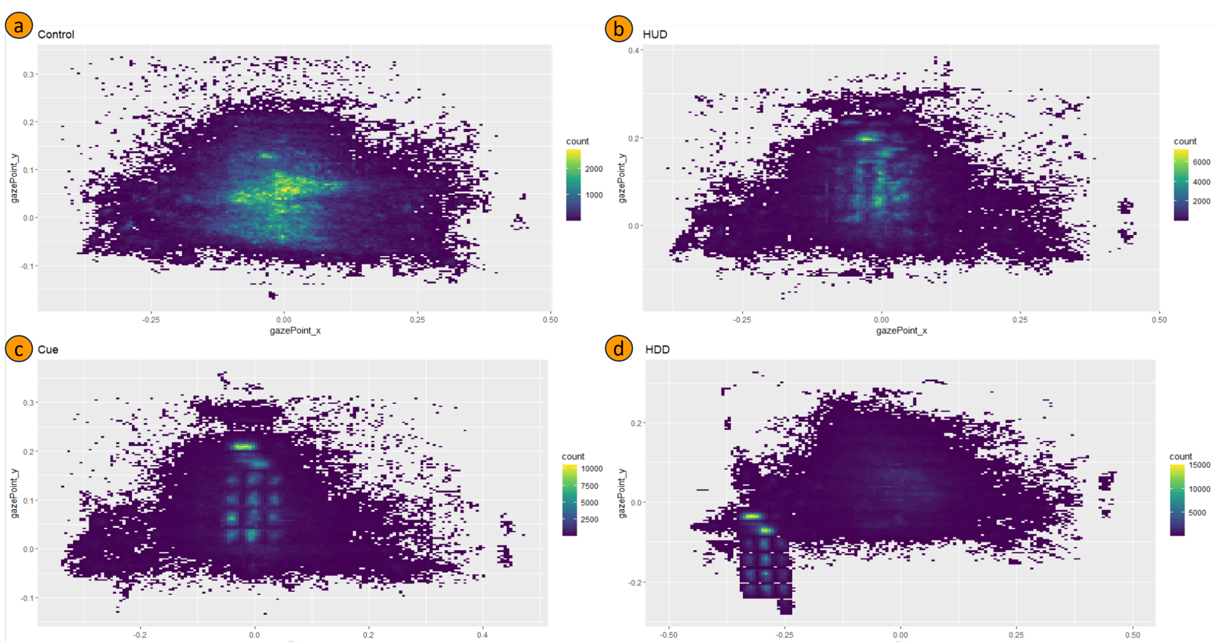


Figure 5.16: Heat maps showing the distribution of fixations in each of the conditions a) Control, b) Static AR HUD, c) Dynamic AR HUD and d) AR HDD. The outline of the keypad NDRT is distinctly visible in c), the Dynamic AR HUD condition, in particular the target prompt at the top of the display. This is in contrast to b) the Static AR HUD, which had the same display but did not move to reveal the road to participants at any point.

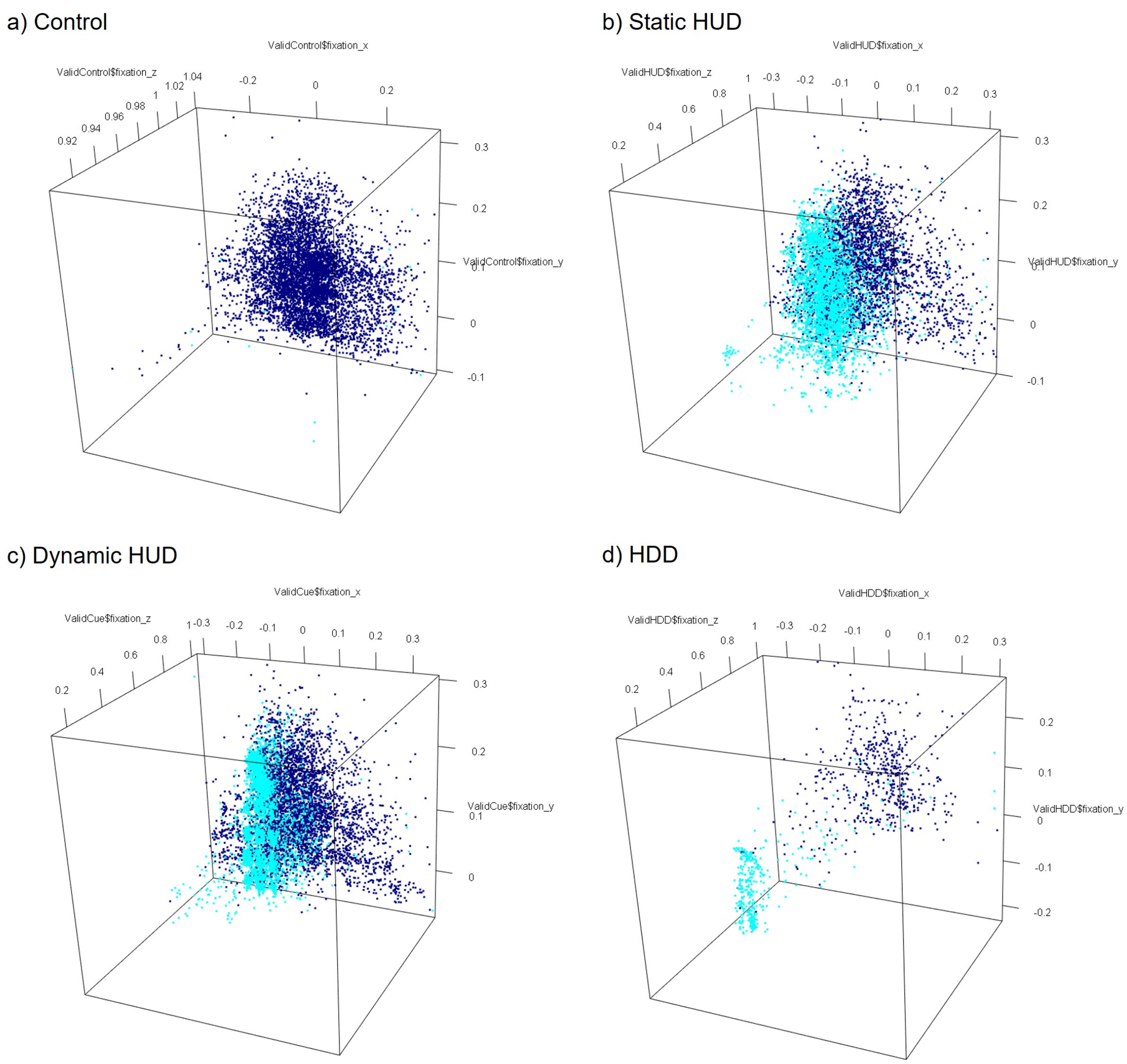


Figure 5.17: 3D Heat maps showing the distribution of fixations on either the Screen (Navy) or AR (Cyan) in each of the conditions a) Control, b) Static AR HUD, c) Dynamic AR HUD and d) AR HDD. The outline of the keypad NDRT is also more visible in c), the Dynamic HUD compared to b) the Static HUD.

5.9 Experiment 4 Discussion

The results of this experiment found that performing an NDRT was detrimental to situational awareness, regardless of the presentation method. This was also reflected in the lower confidence and subjective attention ratings for each of the NDRT presentation conditions, although confidence ratings were significantly higher in the Dynamic AR HUD condition compared to the AR HDD condition. Notably, the benefits of the Cued AR HUD condition in Experiment 3 were lost with a different, more obstructive NDRT and a more disruptive attentional cue. Participants' performance in the NDRT was significantly greater in the HUD conditions than the HDD presentation, although they were not rated as significantly different on the NASA TLX scale.

One potential significant factor was the participant's ability to interact with the AR task. Nine participants reported no experience with AR before this experiment which may have affected their ability to pay attention to the driving task when using the AR NDRTs. Despite having a training session where they were able to practice the task, many participants anecdotally reported struggling to operate the keypad interface using the headset's hand tracking. This task required visual attention to read the target number, digit recall to store and retrieve the target number from memory whilst inputting it plus hand-eye co-ordination to press the button in the AR display. Additionally, the movement and position of the keys in the Dynamic AR HUD condition was dependent on the location of the hazard on screen in relation to the participant's head position. This was done to create a window around the hazard and ensure that it was always visible when the buttons moved. This may have had the opposite effect in participants unfamiliar with AR, where the movement of the keys was more of a distraction to those trying to process these new layouts than an attentional cue. Future designs should balance the attention capturing nature of these cues, so that they do not capture *too much* attention and prove detrimental to situational awareness.

This experiment also measured the difference in eye movement behaviour of drivers engaged with a distracting NDRT presented in AR while they were simultaneously performing a hazard prediction task. The results suggest that driver eye movements were disrupted by engaging in these NDRTs, regardless of AR presentation format, as hazard prediction scores were significantly lower in all conditions compared to just watching the road. However, different gaze patterns were evident for each AR presentation condition, which will be discussed in turn here.

NDRTs disrupt driver eye movements and situational awareness

Engaging with a distracting non-driving task is known to disrupt driver attention compared to when focusing on the road. While this manifested in different ways in these results depending on the way the NDRT was presented, it reinforces how care needs to be taken when introducing NDRTs into automated vehicles. In the HDD condition, the number of fixations that occurred throughout the whole time watching the video clips was significantly lower than both a Static HUD and a Dynamic HUD with an attentional cue, as well as when only watching the road. This is likely due to the increased demand of having to look back and forth between the screen and the AR display, which was presented at the same position as standard in-car centre consoles (see Figure 5.8). The increased switching between looking at the screen and the AR display meant that there was less focus on either task. Fixations were also significantly shorter in the HDD condition compared to both the Control and Static HUD conditions but not the Dynamic HUD condition, suggesting that this increased switching also led to less processing time for each fixation. This pattern was repeated during the Hazard Window, where there were also significantly fewer fixations, in total and on the correct part of the screen, as there was nothing to indicate when participants should swap between the tasks. This, combined with the significantly poorer performance on the Hazard Prediction task, indicates that presenting an NDRT as an HDD has detrimental effects on driver awareness, with little chance for them to gain awareness of the road scene.

A HUD does not intrinsically benefit driver attention

The Static HUD condition saw a significantly higher number of fixations than the HDD condition but also significantly fewer than the Dynamic AR HUD condition. However, there were no differences from the number of fixations in the Control condition. This was also true for fixation duration, where there was also no difference with the Static HUD and the Control condition, but longer fixation durations than both the Dynamic AR HUD and HDD conditions. On the surface, this suggests that eye movements, despite being engaged in an NDRT, were similar to the Control condition without a distraction. However, the important distinction comes from the number of correct fixations during the hazard window, which was significantly lower than the Control condition, along with the significantly poorer hazard prediction performance. This suggests that while overall scanning patterns may not have been affected, the Static HUD drew attention away from relevant hazard-critical information and prevented participants from maintaining awareness of what was happening in the clips. While gaze patterns were significantly different to the HDD condition, the Hazard Prediction scores show that simply presenting an NDRT as a HUD does not confer any additional benefits to situational awareness.

Simply displaying NDRTs in an AR HUD is not enough to facilitate situational awareness and can even be distracting to the driver to a similar level as taking their eyes off the road. Further work must consider the ways in which an AR HUD interrupts driver attention and how this could be mitigated.

Fixation does not necessarily mean attention, even after cueing

Despite the inclusion of an attentional cue in the Dynamic HUD condition leading drivers to fixate on a specific dangerous area of the road, it did not facilitate situational awareness in the form of hazard prediction. There was a significantly higher number of fixations in the Dynamic HUD condition compared to all other conditions, including the Control, both overall and during the Hazard Window. Yet there were no significant differences in the number of correct fixations in the Dynamic HUD condition compared to the Control, suggesting that augmenting the location of the keys in the NDRT to reveal the hazard location seems to have drawn participants' gaze towards the correct area of the screen. However, this did not translate into improved situational awareness, with hazard prediction scores in the Dynamic HUD condition still significantly lower than in the Control condition and in line with the HUD and HDD conditions. Fixations in the Dynamic HUD condition were significantly shorter than all other conditions in the hazard window and shorter than all except the HDD condition overall. Taken with the previous results, this could suggest that participants were making more, shorter fixations in the Dynamic HUD condition, having received a cue that a hazard was imminent, but were not able to sufficiently process the driving scene after switching from the NDRT.

Looking but Failing to See

This pattern of results appears to demonstrate the 'Look But Fail To See' effect in action. The distracting nature of the NDRT did not allow enough time for drivers to switch to and sufficiently process the driving task, despite the fact their gaze was on the hazardous area of the road. This could also be evident in the overall patterns of fixations in Figure 5.16. Qualitatively, the outline of the keypad is more visible in the Dynamic AR HUD condition (panel c in both Figure 5.16 & Figure 5.17). This indicates a higher number of fixations on the individual buttons and, in particular, the target prompt, all of which is less evident in the fixation patterns in the Static HUD condition (Figure 5.16b). There were significantly longer fixations in the Control condition compared to all other conditions, followed by the Static HUD, then HDD, and then the Dynamic HUD conditions. More and longer fixations are typically linked with a greater level of processing for relevant driving information, (Yamaguchi et al., 2019). The longer fixation times and higher fixation counts, paired with the higher Hazard Prediction scores in the Control condition, corroborate this finding.

In follow-up interviews, drivers indicated that their strategy for completing this condition relied on ignoring the road and just focusing on the NDRT. The significantly higher performance in the Dynamic HUD condition compared to both the Static HUD and HDD conditions supports this. Participants were able to solely focus on the AR task until the cue appeared, unlike in the other AR conditions which required continuous switching between the two tasks. Some drivers expressed frustration with the attentional cue as it disrupted their performance on the AR task. Rather than consciously noticing the buttons highlighting a particular area of the road, and deciding to switch task and process the road scene, drivers were too engaged with the NDRT, and the disruption of this drew attention away from the driving task more, even if fixations were on the correct part of the screen to perceive the hazard.

The effects of the HUDs, in particular the Dynamic HUD, on eye movements are significant when considering how drivers might interact with NDRTs in automated vehicles. While a HUD presentation may seem helpful, the results here suggest that it may be of no benefit to drivers to present NDRT contents this way and can be distracting. This mirrors Smith et al. (2015), who highlighted how presenting NDRTs as a HUD does not necessarily result in increased driving performance, even though it was preferred by participants over an HDD. Similar results were found by Schömig et al. (2018), where participants preferred an adaptive AR HUD presenting lane-change information over a static HUD. While these may be a logical way of presenting information to drivers, it appears that with NDRTs, a HUD is similarly as distracting as an HDD. Furthermore, these results shed some light on those of Li et al. (2020). While the same pattern of eye movements is seen here when comparing an HDD to a HUD, the lack of improvement in hazard prediction scores suggests that simply directing driver attention to the road does not necessarily result in increased situational awareness as Li et al. (2020) posit. However, they do recognise that they did not quantify visual attention between the NDRT and the road in this manner. Similarly, Radlmayr et al. (2018) found that horizontal gaze during a visually distracting NDRT was significantly lower than when scanning the road, which led to reduced levels of situational awareness.

Even when the layout of the HUD was augmented in the Dynamic HUD condition to facilitate attention to the hazard, there was still no benefit to hazard prediction scores despite similar behaviour to the Control condition. Pakdamanian et al. (2022) demonstrated that introducing context-dependent cues into a range of NDRTs improved takeover reaction time and performance after regaining control. The results presented here are not concordant with this, where the attentional cue designed did not provide any benefit to hazard prediction. Pakdamanian et al. (2022) also claim there was improved situational awareness with their warning alerts, with an increase in fixations on the road.

However, the results from this experiment found that hazard prediction scores in the Dynamic HUD condition were significantly worse than when watching the road, indicating that driver's situational awareness was impaired by the NDRT, in spite of the inclusion of an attentional cue and an increase of fixations in the correct area of the screen. This is perhaps due to an attentional resource conflict (Wickens, 2002; Horrey and Wickens, 2003), with both driving and NDRT being in the visual domain here, whereas Pakdamanian et al. (2022) used alerts specific to the NDRT modality to draw the attention of drivers to the road. Multimodal alerts have been shown to better convey danger to drivers over unimodal ones (Politis et al., 2014b, 2017). It is possible that including a multimodal alert may improve performance on the hazard prediction task or act as more of a signal to switch to the driving task. However, from these results, it is apparent that using driver attention as a substitute for awareness of the road, at least in the visual domain, may be problematic when designing interfaces for presenting NDRTs.

This has implications for driver monitoring systems, which rely on the assumption that a driver fixating on the road is paying attention and aware of their surroundings. When NDRTs are presented in an AR HUD, a driver monitoring system may assume the driver is paying attention to the road when they are actually interacting with an NDRT and not processing what is happening on the road. Braunagel et al. (2016) demonstrated that eye movement classification is sensitive to the NDRT that is being performed and recommended an adaptive classifier to determine where a driver is fixating. Just monitoring a driver's eye movements alone is not enough to evaluate if they are looking at the road, nor if they are understanding the scene they are fixating on. While the ultimate goal of automation is to eliminate the need for a human driver, they will be required to supervise automated vehicles for some time to come. Designers of NDRTs who want to take advantage of the full benefits of automation should consider the distracting nature of these tasks and be mindful that measuring eye movements is not necessarily the whole story when it comes to keeping drivers in the loop. Incorporating relevant driving-related content into the design of the NDRT may be one potential solution for this (Schroeter and Steinberger, 2016; Wu et al., 2023a), further work is needed to establish the effect of this on SA.

5.10 Chapter 5 Discussion

Summary of Findings

The two studies presented here evaluated whether using AR to present NDRTs can help driver's situational awareness. The results showed that an AR HUD provided no additional benefits over an HDD on Hazard Prediction ability, except when an attentional cue was included into the design of the NDRT. However, this was dependent on the workload of the NDRT, with a more demanding task not showing the benefit of an attentional cue.

These experiments help to answer the research questions in the following ways:

- **RQ1)** *Can drivers maintain situational awareness while engaged with an NDRT?*

Participants were able to predict hazards with an above-chance performance in all the NDRT presentation conditions, but still lower than with full attention on the driving task. Though these results do not rule out NDRTs during supervision, it shows that there is a cost to engaging with an NDRT in terms of awareness. Alternative designs, interactions, and presentation methods should be investigated to lessen the negative effect on awareness and ensure safe takeovers.

- **RQ2)** *Does presenting an NDRT via an AR HUD have benefits for situational awareness over a traditional HDD?*

There were no significant differences between the HUD and HDD conditions in either experiment by themselves. Presenting an NDRT via a HUD alone does not have any significant benefits on hazard prediction performance compared to a HDD. This is perhaps due to the attentional requirement of the NDRT, or occlusion of the road scene preventing similar awareness levels being acquired. This suggests that simply presenting an NDRT via a HUD would not be enough to improve situational awareness

- **RQ3)** *Does including an attentional cue in the AR HUD aid situational awareness?*

Adding a dynamic attentional cue to the AR HUD helped bring attention to the road over the HDD conditions in Experiment 3 , whereas the AR HUD condition itself did not. However, the attentional cue did not prove useful in Experiment 4. The keypad dialling task with the dynamic cue in Experiment 4 was rated as having a higher workload than a gem-popping game with a red gem cue in Experiment 3, which may explain the disparity in results. Eye tracking data show that fixations were successfully drawn to the area cued by the AR HUD, but this did not translate into improved situational awareness. Together, these results suggests that attentional cues may provide benefit to situational awareness, but what constitutes an effective cue requires further research.

Limitations and Recommendations for Future Work

The Hazard Prediction task here is used as an approximation for a driver monitoring the road and predicting what happens next, indicative of the *projection* stage of Endsley (1995a)'s situational awareness model. The Hazard Prediction task presented here is a valid SAGAT variation which has had success in discriminating between novice and experienced drivers (Crundall, 2016; Ventsislavova and Crundall, 2018) This is not entirely representative of the driving task as a whole, as it requires only a short period of attention in a controlled setting for the duration of the clip, rather than sustained attention over a longer period of time. As Mackenzie and Harris (2015) point out, driver eye movements can differ between passive and active driving. This is not necessarily the technique used in longer driving scenarios where supervision is more likely to occur, i.e., motorways. Here, factors such as fatigue, distraction or the length of the drive are likely to affect attention to the road (Kee et al., 2010; Ting et al., 2008). Though the hazard prediction test has merit as a measure of a driver situational awareness (Crundall, 2016), future research should compare the effect of engaging with an NDRT in a HUD vs an HDD in a more realistic road scenario, i.e., an automated driving simulator or as a passenger in an on-road study, to measure the behaviour of drivers taking control after having been distracted.

Similarly, whilst footage was collected from a wide range of road types and environments, the majority of hazard clips occurred in close urban or suburban roads. This is consistent with areas where most road collisions occur (UK Government, 2022a), so this is likely where TORs would be common, as well as more driving in general. However, this task cannot measure a driver's predictive ability for more novel or environmental hazards that are difficult to predict, such as a patch of ice on the road or an oncoming vehicle obscured by a bend. Navigating these hazards relies more on faster reactions and attention to the road rather than predicting what is about to happen, as they are unpredictable.

Additionally, the stimuli presented in both studies were visual. Politis et al. (2014b) have shown that multimodal interfaces are much more effective at attracting driver attention in the event of a TOR (Politis et al., 2014b, 2017). Pakdamanian et al. (2022) also showed that cueing attention multimodally leads to increased situational awareness and safer management of takeovers, and Ma et al. (2023) found different neural activation patterns for unimodal and multimodal interfaces, though there were no differences in workload for locating notifications. It has been shown that performing tasks that compete for the same resources simultaneously results in poorer performance (Brünken et al., 2002; Horrey and Wickens, 2003), something which is also apparent for in-car NDRT inputs (Roider et al., 2017). This may explain the lack of effectiveness of the attentional cues compared to baseline performance in both studies.

NDRTs with lower attentional requirements or that require different cognitive resources such as voice interfaces or auditory tasks may receive a greater benefit from attentional cues. Further research should compare NDRTs using different modalities and how conflicting versus compatible modalities moderate the effectiveness of attentional cues. Furthermore, future work should also investigate whether an AR HUD would benefit from multimodal cues, or if a multimodal NDRT has a greater impact on situational awareness. Finally, follow-up research into how drivers pay attention to attentional cues during an extended automated drive would also reveal how they can be used when a TOR is not needed.

Previous research has shown that the accuracy of the HoloLens 2 is not as reliable as specific eye trackers. Aziz and Komogortsev (2022) highlighted calibration issues when using the HoloLens 2, which results in a certain amount of drift over time. Unfortunately, the experiment presented here was conducted before this experiment was conducted, and so their recommendations were not followed. As such, the analysis technique used here has been limited to using the much broader I-AOI classification technique. While precise fixation locations cannot be reported, there is still value in using these AOI classifications to show differences in eye movements between AR presentation conditions.

Similarly, most eye-tracking research collects data at 60Hz, with the 'standard' being considered 100Hz. Using the HoloLens 2, it was not possible to collect measurements at this frequency due to the headset limit of 30Hz. However, the ability to measure eye movement behaviour when also using an AR headset presenting 3D holograms is not easily implemented using more traditional eye-tracking hardware. Whilst the measurements here might not be at the same frequency as dedicated eye-trackers, the data collected allow the viewing of realistic interactions with an AR interface.

Finally, both the Hazard Prediction and AR NDRTs were presented in the visual domain, increasing the overlap between attention systems (Wickens, 2002; Horrey and Wickens, 2003). Schartmüller et al. (2021) found that presenting an NDRT via a HUD significantly increased perceived workload and had a negative impact on gaze behaviour, compared to an audio interface. Further research should also investigate how engaging with different multimodal NDRT interfaces might affect a driver's situational awareness. For instance, whether spatial audio (Beattie et al., 2014; Wang et al., 2017b), haptic (Telpaz et al., 2015; Di Campli San Vito et al., 2020a), or thermal cues (Di Campli San Vito et al., 2018) might have improved the driver's awareness of the road environment (Geitner et al., 2019). However, given the projected popularity of in-car AR, there is merit in investigating the specific effect of engaging with a distracting visual interface that overlaps with the predominantly visual driving task.

AR HUDs: Help or Hindrance for NDRTs?

The benefit of using HUDs to display information is not a new finding (Medenica et al., 2011; Smith et al., 2015; Jose et al., 2016), but there has been little research into how presenting a *non-driving related task* affects the underlying hazard awareness of the driver. While most AR HUD concepts tend to focus on displaying driving-relevant information to the driver (D'Arcy, 2022; Firth, 2019a), people report wanting to engage with NDRTs in AVs (Panagiotopoulos and Dimitrakopoulos, 2018). The time for a driver to regain situational awareness is around 10 seconds (Walch et al., 2017), which is increased when using a handheld device for an NDRT (Zhang et al., 2019) as well as when engaged with an NDRT presented in a HUD (Li et al., 2020). The results from the studies presented here corroborate the findings that any heads-down activity impairs the driver's ability to maintain awareness of the road (Liu and Wen, 2004b). However, they also indicate that, unlike driving-related information, engaging with an NDRT via a HUD also hinders driver awareness. This is important since current regulations and guidelines allow drivers of AVs to engage with NDRTs such as playing games or watching movies within the car (BBC, 2022), as long as they are able to regain control of the vehicle if needed (UK and Scottish Law Commissions, 2020). To maintain driver awareness whilst engaged with an NDRT, simply switching from an HDD to a HUD does not provide the benefits seen with driving-related information (Lindemann et al., 2018; Jing et al., 2022). Design considerations must be made to account for the distracting nature of NDRTs.

Schömig and Metz (2013) have shown that drivers can engage with NDRTs in a situationally aware way, and previous studies have investigated the feasibility of AR interfaces for NDRTs during the supervision (Riegler et al., 2019b, 2022). However, these are not the same results as found here, where AR HUDs provided no benefit to hazard awareness over their HDD counterparts, with all NDRT conditions suffering worse hazard prediction performance than full attention. This is similar to Radlmayr et al. (2018)'s findings that presenting a balloon-popping game via a HUD led to poor performance on a SAGAT test compared to no secondary task. A distracting AR HUD could lead to inappropriate or delayed reactions to a critical TOR based on poor situational awareness. Even though AR content was overlaid onto the driving task, the 'Look But Fail To See' phenomenon (Wolfe et al., 2022) persisted. While the NDRT in Experiment 3 was designed to have a relatively low impact on awareness, the constant vigilance for non-driving content may have been one cause for the decrease in performance. Furthermore, the NDRT and the Hazard Prediction task were both visual. This is concordant with previous work showing that the workload of different NDRTs affected TOR performance (Yoon and Ji, 2019; Müller et al., 2021). When driver attention is not on the driving task, the benefits of an AR HUD may be lost if the task is displayed at eye level without design changes.

Modifying the NDRT to take advantage of the dynamic spatial and visual aspects the display offers however can be used to aid situational awareness. Pakdamanian et al. (2022) found that providing contextual awareness notifications for different NDRT modalities led to increased situational awareness and smoother takeover requests. Jiang et al. (2023) also found that certain types of "situational" mobile games increased situational awareness compared to more distracting games. These studies suggest that providing contextual attentional information to drivers when they are engaged in an NDRT to aid awareness could be possible. The attentional cues presented in Experiment 3 were a first attempt to measure this regarding presenting an NDRT via an AR HUD. Concepts exploring how this could be applied to a gamified HUD involve displaying game elements over important road features (Schroeter et al., 2014; Togwell et al., 2022). However, this effect did not carry over to Experiment 4. The difference here being that the attentional cue was more visually distracting and the NDRT obscured the road more, leading to a greater perceived workload. This suggests that the attentional cue's design is important. A distracting cue which draws more attention than the danger it is supposed to be cueing will have the opposite effect on driver awareness, which is one of the biggest challenges for realising in-car AR (Riegler et al., 2020).

Implications for Designs of AR HUDs

The use of HUDS in AVs for non-driving related tasks provide a greater challenge than simply presenting information, as they create competition for driver attention. The results from the experiments described here suggest that simply presenting an NDRT in a HUD is not enough if drivers are still required to be aware of the road. Current discussions around responsibility and awareness between humans and AVs mostly assumes a binary relationship with either the human (User-in-Charge) or the computer (No-User-In-Charge) in control (UK and Scottish Law Commissions, 2020). However, it is likely that driver awareness will be impaired following a system-initiated TOR if attending to an NDRT (Janssen et al., 2019). A more comprehensive approach would be to model the relationship as a continuum of responsibility that shifts between the driver and the AV throughout the drive (Marcano et al., 2020). Janssen et al. (2019) set out a framework that highlights different stages of disengaging and re-engaging with an NDRT during an automated drive, and how driver attention progresses through these stages, rather than instantly switching task from non-driving to driving task. Importantly, they point out how driver awareness is likely to be impaired if a binary transfer of control is enacted without sufficient prior warning.

Whilst the role of driver lessens as the capabilities of AVs increase, full automation is still many years away. Hazard prediction is one of the key challenges for reaching L4 AVs (Schwalb, 2021; Mallozzi et al., 2019). L3 vehicles, however, are now being sold in the USA and Germany (Harley, 2023), and the UK Government is creating legislation defining who is responsible in an AV both with and without a driver able to take control (UK Parliament, 2023). Issues with implementation has resulted in a slower rollout of L4 vehicles than anticipated (Gordon, 2023). L3 vehicles are therefore a more realistic goal for automated driving, which can use driver expertise in hazard prediction as a failsafe. However, systems which can keep drivers attending to the road without negating the benefits that automation can bring are needed. A dynamic HUD which allows drivers to engage with NDRTs and also preserves their awareness of the road through attentional cues is one solution to this issue. For example, with the AV in control, the driver is free to engage in NDRTs and take their attention off the road. As the AV starts to lose confidence in its ability to predict a complex road scene, or its ability to safely navigate, it could start to notify the driver by modifying the NDRT to include information about the road, e.g., the colours surrounding a HUD displaying emails start to change (Yang et al., 2018), or the elements of a game change to reflect objects on the road (Wu et al., 2023a). If the AV is not able to resolve these issues itself, it can then start to notify the driver that they should be prepared for a TOR. Rather than being expected to assume absolute control with no awareness of the road environment, a driver being shown these dynamic cues would be able to take on more responsibility for the driving task without needing to spend time reacquainting themselves with the road. In extreme cases, the driver would be able to safely take control of the vehicle rather being thrust into a dangerous driving scene and have to make a decision based on poor awareness.

This all needs to be considered with the ultimate aim of still allowing drivers to engage with NDRTs, so as not to lose the benefits of automation. Previous research shows how novelty could prove more attention capturing than the cue itself (Itti and Baldi, 2005; Ernst et al., 2020), which may be an explanation for the results in Experiment 4. A display which, upon detecting a hazard, opens up to reveal the road beneath may allow drivers full view of the hazard, but the disruption to the task actually draws attention away from the road. The balance between creating an attentional cue which provides important information but does not capture *too* much attention requires further research if attentional cues are to be incorporated into NDRTs.

Recommendations

The results from these two studies have implications for designers of in-car infotainment systems that can be used for NDRTs during automated driving. Whilst currently confined to a heads-down internal screen, the progression to higher levels of driving automation along with an increase of in-car mixed reality displays will allow more sophisticated interfaces to be implemented. The results from the two studies presented here indicate that simply presenting non-driving content in a HUD does not provide significant benefits to driver awareness by itself.

From this are a list of recommendations for designers of in-car infotainment displays in order to limit the impact they may have on driver awareness and take advantage of the dynamic nature that mixed reality interfaces can provide:

- **Include dynamic attentional cues which draw attention to the road**

  Simply displaying NDRTs in an AR HUD is not enough to facilitate situational awareness, and can even be distracting to the driver. NDRTs should include a dynamic cue that brings attention to specific areas of the road.

- **Utilise change in colour and positioning**

  Giving attentional cues a distinctive appearance that separates them from other NDRT elements and enhances their ability to capture attention. Creating a visually attention-capturing cue requires it to be visually distinct from the underlying task.

- **Do not disrupt performance of the NDRT**

  A dynamic cue which violates user assumptions of how to perform the NDRT can be more distracting, and so the design of dynamic attentional cues should consider the design of the NDRT and fit into how users typically engage with and what they expect of the task.

- **Consider the overall workload of the non-driving task**

  A more demanding task requires more attentional resources, which leads to less attention available to monitor the road. The increased workload of tasks such as reading text or inputting information needs to be considered when trying to draw driver attention back to the road with an attentional cue.

5.11 Conclusions

This chapter presented two experiments which compared the impact of presenting a NDRT in a HUD or a HDD on a Hazard Prediction task. In Experiment 3, performance on this task was significantly impaired compared to a baseline without any non-driving task as a distraction. Presentation in an AR HUD including an attentional cue did show significantly less impact than HDD conditions. However, performance was still poorer compared to when participants completed the task without a distracting NDRT. Experiment 4 expanded on these results by using a more realistic non-driving task including an attentional cue, with a design informed from an expert focus group. Participants' performance on this task was significantly impaired compared to baseline without any non-driving task as a distraction, regardless of presentation method or the inclusion of the attentional cue.

The results from these experiments help to address **RQ1** of the thesis (section 1.3), demonstrating that the demanding nature of a NDRT also impairs situational awareness as it impaired reaction time, as seen in Chapter 3. Furthermore, in answer to **RQ2** of the thesis, presenting the AR display in a HUD location did not have the beneficial effects that are seen in the literature for driving-related information. Simply presenting the non-driving tasks via an AR HUD to keep their eyes on the road may not be the solution to keeping drivers in the loop. Analysis of gaze behaviour showed that, despite fixating on the road, a HUD presentation provided no additional benefit to hazard prediction.

Furthermore, to address **RQ3** of the thesis, a Dynamic HUD with an attentional cue that successfully drew the driver's attention to the hazardous area of the road did not translate into improved situational awareness. When assessing the situational awareness of drivers, simply using their gaze does not give the full picture. The results from this chapter have implications for the use of driver monitoring systems, which evaluate driver attention if their eyes are on the road. They suggest that these systems risk overestimating the ability of a driver to maintain awareness of the road, especially if they are engaged with a distracting NDRT. Designers of in-car displays should consider these results when creating in-car displays and consider the workload of the NDRTs that drivers might engage with. This poses the question of whether it is possible to use an attentional cue within an AR display that can maintain situational awareness during a NDRT at the same level as without a distraction. The next chapter explores this further, expanding on **RQ3** of the thesis.

Chapter 6

The effects of using a Virtual Agent to Signal Danger on the Hazard Prediction ability in Distracted Drivers

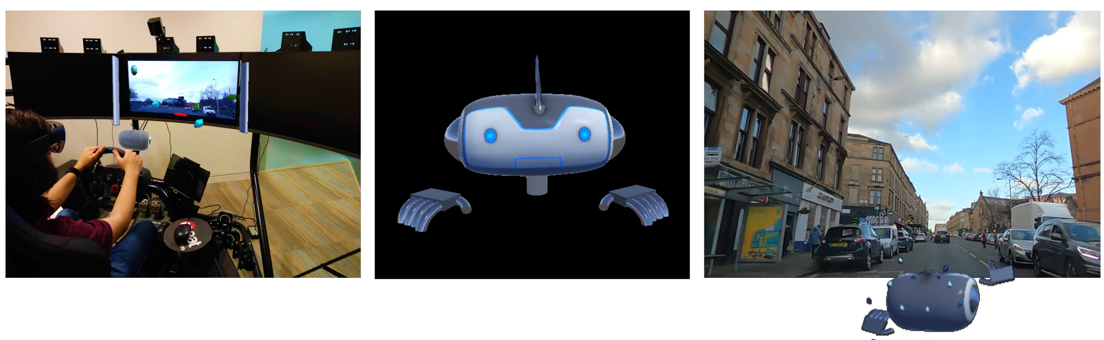


Figure 6.1: View of the experimental setup of the driving simulator with an example of the augmented reality task (left), the front on view of the virtual agent used (centre) and an example of the animated social cue with the agent pointing towards the hazard in the clip (right).

While in-vehicle interfaces can benefit from being characterised as social interactions (e.g., conversational agents providing navigational instructions), displaying safety-critical information to drivers through social cues has not been evaluated. This chapter investigates whether social information from a virtual agent can aid driver awareness during a distracting non-driving related task (NDRT). Two experiments evaluated the hazard prediction ability of drivers performing NDRTs with either a visual cue or a virtual agent indicating danger. Results demonstrated that when an attention-capturing element was included, the head-turning of the virtual agent was as beneficial as a visual cue. Furthermore, displaying these cues during an NDRT maintained hazard prediction performance as when watching the road without distraction. This chapter addresses **RQ3** for the thesis, and has implications for using social information in attentional displays to aid driver awareness.

6.1 General Introduction

Operation of a vehicle usually falls to a single driver, who takes on full responsibility for manually controlling the vehicle and attending to the road. As automated vehicles become more advanced, the responsibility for these driving tasks will be increasingly shared with a computer. Consequently, drivers can focus on Non-Driving Related Tasks - (NDRTs) while the vehicle operates autonomously. Unfortunately, despite the promise of fully replacing drivers with computers, the progression of automated vehicle technology has not matched the enthusiasm for its benefits (Hancock, 2019). As such, current legislation requires the driver to remain attentive to the road and aware of the vehicle's state at all times (UK Parliament, 2024; UK and Scottish Law Commissions, 2020), regardless of what tasks the vehicle can perform. This creates competition for the driver's attention between maintaining awareness of the road and the NDRT. Distracted drivers show reduced driving performance after taking control (Dogan et al., 2019; Yoon and Ji, 2019) and cannot process the road effectively (Goodge et al., 2024). To mitigate this, new methods of displaying information to drivers must be developed to aid their awareness of the road while distracted by an NDRT.

Using Augmented Reality - (AR) to display driving-related information benefits attention when the driver controls the vehicle (Karatas et al., 2020; Wang et al., 2022b). Presenting information at eye-level with the road as an AR Heads-Up Display - (HUD) has seen benefits in automated vehicles for reducing reaction times to take over requests (Wintersberger et al., 2018b) and crash warnings (Kim et al., 2013). AR HUDs have also been found to facilitate smoother transitions between manual and automated driving (Langlois and Soualmi, 2016) and increase trust in automation (Wintersberger et al., 2018a; Oliveira et al., 2020). However, questions remain over the impact of presenting a distracting NDRT via an AR HUD on the ability of drivers to anticipate danger, a key aspect of their situational awareness of the road (Crundall, 2016; Endsley, 1995a). Regaining awareness after resuming control is impaired when performing an NDRT on a handheld device (Zhang et al., 2019) and in a HUD (Li et al., 2020). While AR HUDs have shown benefits in highlighting danger (Wang et al., 2022b; Schömig et al., 2018), awareness is impaired if the driver is also performing an NDRT (Hungund and Pradhan, 2023; Radlmayr et al., 2018), even when attention is cued towards danger (Goodge et al., 2024). Overlapping task demands of a visual NDRT on the visual driving awareness task may be one cause for this (Wickens, 2002; Wandtner et al., 2018), but it is still unclear how to use AR to inform drivers of hazardous road events while they are distracted by an NDRT.

Incorporating virtual assistants to communicate information to drivers may provide benefits over current visual or auditory interfaces (Wallbridge et al., 2022; Wang et al., 2022a; Karatas et al., 2018). Humans can process social information (behavioural cues relevant to social interaction) at faster rates than visual information (Thornton and Conway, 2013; Frischen et al., 2007). Social cues such as head-turning behaviour, where humans orient their attention towards where another person is pointing their head (Langton et al., 2000; Rubio-Fernandez et al., 2022), has been shown to effectively communicate location in human-robot interaction settings (Yamazaki et al., 2008; Hoque et al., 2011; Sheikhi and Odobez, 2015). Current research investigating virtual assistants in vehicles has seen potential benefits in reducing perceived workload (Karatas et al., 2015, 2016), increasing perceived safety (Karatas et al., 2018) and increasing trust in automation (Wallbridge et al., 2022). Using social cues to display information about the road may provide additional benefits in aiding driver awareness of the road while they are distracted by an NDRT, compared to a visual cue alone.

This chapter investigates whether a virtual assistant exhibiting social cues highlighting danger can benefit driver situational awareness. Specifically, can the head-turning behaviour of a virtual agent displayed in AR aid hazard prediction, compared to a traditional visual cue. Two experiments are presented that investigate how prediction ability is affected by including an attentional cue, either as a visual or social modality, while drivers are engaged with a visually distracting NDRT. In the first experiment, the hazard prediction performance of drivers engaging with an NDRT that included a cue that only highlighted the position of the hazard was measured. Performance with either a visual cue or an agent cue included in the NDRT was compared to performance both without any cue included and without an NDRT. The results showed performance in all conditions with an NDRT was significantly poorer than with no NDRT, regardless of the type of cue included. In the second experiment, hazard prediction performance was again compared between cues indicating the hazard's presence and position. A colour-changing visual cue, a colour-changing agent and an animated social agent, which all alerted the driver to the presence of a hazard and highlighted its location, were compared to hazard prediction performance without an NDRT. The results from this experiment found that both the colour-changing visual and agent cues maintained the same level of performance as when there was no NDRT, but the animated social cue condition was significantly poorer than all other conditions.

6.2 Related Work

Automated vehicles will allow drivers to take their attention off the road onto NDRTs (Panagiotopoulos and Dimitrakopoulos, 2018). While the ultimate goal is a fully autonomous vehicle that requires no human intervention, achieving this at scale has proved challenging (Kosuru and Venkitaraman, 2023; Tengilimoglu et al., 2023). Therefore, governments worldwide are starting to prepare for partially automated vehicles, where a driver is still required to pay attention to the road during an automated drive (UK and Scottish Law Commissions, 2020; UK Parliament, 2024; The European Parliament, 2022). Should automation fail, the current safeguard is for the vehicle to issue a Takeover Request - (TOR), an alert signalling the driver to resume full control of the vehicle. These TORs are designed to capture the driver's attention and alert them that they need to take, or are about to be given control. While this is typically measured through driver reaction time, the usefulness of the TOR paradigm may be overstated (de Winter et al., 2021). It relies on the assumption that a driver is fully aware of the road scene at all times during an automated drive and able to smoothly regain control of the vehicle to deal with the current situation after a TOR has been issued. This is not plausible when a driver is engaged with a distracting NDRT which takes their attention away from the road. Much research has shown that distraction reduces driving performance (Papantoniou et al., 2017; Oviedo-Trespalacios et al., 2016). Takeover performance is inhibited after drivers have been engaged with NDRTs (Weaver and DeLucia, 2022; Soares et al., 2021; Radlmayr et al., 2018), with reaction time to TORs (Zeeb et al., 2016; Wu et al., 2019), hazard perception performance (Zangi et al., 2022) and driving eye-movement patterns (Li et al., 2020; Wu et al., 2022, 2021) all being disrupted. An unexpected TOR results in a period where the driver is required to switch tasks from the NDRT to controlling the vehicle (Janssen et al., 2019), during which driving performance is impaired (Soares et al., 2021; Wu et al., 2022; Eriksson and Stanton, 2017).

Guiding Driver Attention

To try and improve the effectiveness of TORs, attempts have been made to communicate the vehicle's state and information about the road to the driver during an automated drive without drawing their attention away from the road. Capallera et al. (2022) highlight how this has involved creating peripheral interactions so as not to intrude on the driver's view of the road. Examples include LED strips on the dashboard (Yang et al., 2018; Löcken et al., 2016), ambient lights around the steering wheel (Mok et al., 2017; Matviienko et al., 2016), combined dashboard and centre console displays (Wang et al., 2017a) and peripheral lights on a pair of glasses (Van Veen et al., 2017).

Further research also demonstrates the use of vibration or thermal cues to provide navigational cues (Di Campli San Vito et al., 2019) or signal a TOR (Di Campli San Vito et al., 2020b) to drivers in a less intrusive manner. These interfaces provide information while the driver is attending to the road or manually controlling the vehicle. However, the ability to perceive this information is impaired for drivers distracted by an NDRT (Zhang et al., 2019; Zangi et al., 2022). Studies investigating drivers resuming control of an automated vehicle after engagement with an NDRT suggest that reaction times (Dogan et al., 2019) and braking responses are impacted (Weaver and DeLucia, 2022; Wandtner et al., 2018). Using AR for in-vehicle mixed reality displays (McGill et al., 2020, 2022) has been proposed as an efficient way to present information to drivers. Displaying information at eye level in this way as a HUD has benefits for driver attention. A meta-review by Niu et al. (2024) highlights how warnings presented in a HUD improve reaction times compared to HDDs. Furthermore, AR HUDS can display more dynamic and detailed information to drivers than traditional HUDs that only display static information (Gabbard et al., 2014; Riegler et al., 2019a). This has benefits for highlighting pedestrians (Wang et al., 2022b), road crossings (Karatas et al., 2020) and communicating planned manoeuvres (Schömig et al., 2018). AR HUDs have also been shown to reduce distraction (Jing et al., 2022), increase recognition of vulnerable road users (Rusch et al., 2013), and aid lane change manoeuvres (Langlois and Soualmi, 2016). These interfaces draw attention towards important aspects of the road to aid a driver's situational awareness; their mental representation of the current road scene (Endsley, 1995b). For example, Dynamic AR HUDs can inform drivers of upcoming road events in real-time (Schömig et al., 2018) or cue attention to important or dangerous road areas (Lindemann et al., 2018). However, AR HUDs are typically evaluated when the display presents driving-related information, which aims to improve the driver's awareness of the road. One of the main uses of an AR HUD would be to present NDRTs to drivers (McGill et al., 2020, 2022), yet Niu et al. (2024) note that few studies investigate the effects of using *AR* HUDs for engaging with NDRTs. What has been done indicates that presenting an NDRT as a HUD display does not necessarily convey the same benefits (Smith et al., 2015; Schömig et al., 2018; Goodge et al., 2024). Designing an AR interface to provide driving-related information to a driver distracted by an NDRT requires consideration of how best to communicate information the driver perceives and understands.

Shared Attention to the Road

In partially automated vehicles, sharing responsibility for the driving task between humans and computer has been characterised as "co-driving" (Wu et al., 2023b). During partially-automated driving, the human driver still needs to be informed of the current state of the vehicle, the road environment and the vehicle's understanding of the road scene.

Studies have shown that having a driving-experienced passenger in the vehicle can benefit a driver's awareness and response times to danger (Chandrasekaran et al., 2019). Similarly, while mobile phone conversations impair driving performance, (Crundall et al., 2005), in-person conversations do not have the same impact, attributed to both driver and passengers moderating the extent of their conversation based on the current demands of the driving scenario (Drews et al., 2008). This includes changing the conversation topic to be about the road, aiding situational awareness and navigation (Drews et al., 2008). Though driving is largely an individual task, this suggests that drivers can benefit from information from another perspective.

While previous work focuses on verbal interactions, incorporating mixed reality interfaces into vehicles could display a passenger's (or an automated vehicle's) understanding of the road in novel and informative formats, going beyond speech. Maurer et al. (2014) proposed a prototype for displaying passenger gaze behaviour to the driver of a manual vehicle using mixed reality. In a driving simulator setup, a yellow circle was displayed on the screen to show the driver where their passenger was looking to aid navigation. Drivers reported that using this interface felt easier than verbal communication for directions. This concept was evaluated further by Trösterer et al. (2015), who compared this passenger gaze display with verbal communication or an idle driver. They found that drivers collaborating with a co-driver found the task less stressful and reduced perceived workload. However, they found mixed results regarding the gaze display, with no benefit to driving performance. A follow-up study utilised an LED strip mounted on the dashboard to display passenger gaze (Trösterer et al., 2019). Drivers were tasked to drive along a pre-defined route while a passenger aided navigation using their gaze, fixating on objects in the road scene for the driver to identify. A correct object identification rate of 77% was observed (higher for identifying vulnerable road users: 85%), with drivers reporting the gaze visualisation was useful for aiding navigation (Trösterer et al., 2019). Similar studies evaluating AR HUDs, which guided eye movements towards the road helped to reduce detection time (Fathiazar et al., 2023; Karatas et al., 2020) and increase object recognition (Wang et al., 2022b).

Despite relaying information from another perspective, these displays remain within the visual domain to communicate information to the driver. Within a HUD presentation, displaying visual information is likely to compete for attentional resources with a visual NDRT and has been found to reduce situational awareness (Goodge et al., 2024; Radlmayr et al., 2018). Trösterer et al. (2015) reported that, despite the benefits, some drivers reported the constant passenger gaze display as intrusive and distracting. This is likely, according to Wickens (2002) Multiple Resource Theory, due to the concurrent driving and non-driving tasks competing for the same attentional resources, leading to reduced performance in both tasks (Head and Helton, 2014; Fereydooni et al., 2019; Ward and

Helton, 2022; Jackson et al., 2023). This raises an issue when considering how to display road information to drivers so that it does not overlap with an NDRT. For example, visual alerts indicating danger could conflict with a visual display showing a video, and audio alerts can blend into music or background road noise (Šabić et al., 2019). Providing cues relevant to the context of the NDRT has been shown to benefit takeover reaction times (Pakdamanian et al., 2022), but the driver's awareness of the road scene after a TOR has not been evaluated before. Designing an effective attentional cue that provides this information to distracted drivers remains a challenge, particularly creating an attentional cue which does not impair on the driver's ability to process the road.

Social interaction with(in) the vehicle

A novel way to communicate information could be using *social information*; behavioural cues such as speech or eye gaze which humans use to facilitate communication (Frith and Frith, 2007). Humans are more efficient at processing social information compared to visual information (Thornton and Conway, 2013; Frischen et al., 2007), yet there has not typically been a place for social information used for in-vehicle interfaces. Previous work mostly considers social interactions between the driver and other road users (Dey et al., 2019; Walch et al., 2019; Ezzati Amini et al., 2021). However, incorporating information that forms part of social interaction is a potential solution to creating attentional cues which do not intrude on the driving task. For example, using speech warnings for alerts improved response times (Politis et al., 2015) and driving behaviour in critical situations (Politis et al., 2014a) over audio-based cues. Beyond this, characterising a speech interface as a social interaction with an in-vehicle virtual assistant has been shown to benefit drivers of automated vehicles, reducing driver fatigue and increasing alertness (Huang et al., 2024).

Joint Attention

While these studies employed speech interfaces to improve driver awareness, they only encompass one aspect of social information: *speech*. Yet, the capacity for humans to process social information goes beyond speech alone; eye gaze and head turning are useful social cues to work out where another person is attending to (Kluttz et al., 2009), known as establishing *Joint Attention* (Moore et al., 2014). Using these social cues is cognitively efficient (Herrmann et al., 2007; Thornton and Conway, 2013), with people responding quicker to a head turn than a visual cue. When someone turns their head, our instinct is also to look and see what they have noticed (Langton et al., 2000), which is used as an inference for gaze (Frischen et al., 2007; Rubio-Fernandez et al., 2022) in the absence of eye contact.

Consequently, head-turning behaviour has been used in human-robot interaction research to create a shared point of focus, since eye gaze can be difficult to establish in robots (Sheikhi and Odobez, 2015). Having a robot exhibit human-like head turning and gaze behaviours has seen benefits in recalling a story told by a robot (Mutlu et al., 2006) and increasing likeability and anthropomorphism in human-robot handovers (Kshirsagar et al., 2020). Directing a robot's gaze towards an object is also an effective way of drawing a person's attention to that object (Admoni and Scassellati, 2017; Hoque et al., 2011). This has seen specific applications in settings such a robot museum guide (Kuno et al., 2007; Yamazaki et al., 2008; Sheikhi and Odobez, 2015) with participants successfully identifying exhibits that a robot turns its head towards.

Robots in Robot Cars

In a driving context, similar work has found that introducing robots as an embodiment of a virtual assistant can benefit driver attention. Karatas et al. (2015) introduced an in-car multiparty robot called NAMIDA, consisting of a dashboard-mounted robot with three spherical heads that were able to rotate and direct attention towards the road. When the robot was engaged in multiparty conversation, Karatas et al. (2015) found lower perceived attentional, visual and auditory demand of drivers. They also found interacting with NAMIDA encouraged more fixations on the road (Karatas et al., 2016). When NAMIDA directed their gaze towards dangerous road events, ratings of agency and safety were increased (Karatas et al., 2018). In a separate study using the Nao robot, Wang et al. (2022a) found that scores on a situational awareness test were higher when drivers were accompanied by a robot providing either informative or conversational interactions, compared to speech commands alone. Furthermore, Wang et al. (2022b) also found that embodying a virtual assistant as a robot increased ratings of competence and reduced driver perceived workload.

This work suggests there are benefits to incorporating an embodied virtual assistant as an in-vehicle interface and characterising interaction with a vehicle in a social capacity. However, installing a robot into every future automated vehicle is likely to be complex, expensive and undesirable to some drivers (Tussyadiah et al., 2020; Nachmann et al., 2020). Additionally, there is evidence that drivers prefer a conversational-only agent compared to a fully embodied robot (Dong et al., 2020). With more advanced in-vehicle interfaces however, (Robertson, 2024; Tisshaw, 2020) and incorporation of mixed reality into the vehicle (McGill et al., 2020, 2022), it is possible to display a greater variety of dynamic information to drivers. The introduction of mixed reality as in-vehicle interfaces allows more sophisticated displays that provide information to drivers *e.g., through an embodied agent.*

Though confined to a standard centre console display, NVIDIA's DRIVE Chauffeur concept which includes their "Omniverse" Avatar demonstrates a move to characterise interaction with an automated vehicle as social, through a conversational agent (Shapiro, 2021; NVIDIA, 2021). Wallbridge et al. (2022) (unpublished) describe preliminary results evaluating the effect of a social agent presented in mixed reality benefited trust and attribution of blame towards an automated vehicle. However, it is still unclear whether social agents can be used as specific attentional aids towards hazardous events in the road. Though Karatas et al. (2018) found that gaze-directing robots increased ratings of safety, they did not measure the effect these had on the drivers gaze behaviour, or their awareness of hazards. However, Tamura et al. (2021) found that the same gaze-directing robots did aid reaction time to congruent visual cues in a laboratory setting. Wang et al. (2022a) showed that an in-vehicle assistant providing information about the road aided situational awareness, though this was only probed while the driver was focused on the driving task and not distracted by an NDRT. Given our ability to process social information more efficiently than visual information, there may be benefits to displaying important cues about the road via a virtual agent as a social cue.

Summary and Research Questions

As new methods of interaction between the driver and the vehicle become available, new modalities for presenting information to drivers are possible. Utilising social information that humans process more efficiently may provide benefits to driver attention above traditional visual displays and aid drivers who are distracted by an NDRT. However, questions remain over how using these social cues can affect situational awareness of drivers. Moreover, is social information inherently easier to process than visual information, or does the distracting nature of an NDRT mean that specific cueing of attention is still required? Based on concepts from driver attention and human-robot interaction literature, this chapter sets out to examine:

- **Ch6 RQ1** Is hazard prediction performance of a distracted driver affected by including visual or social cues showing positional information alone?
- **Ch6 RQ2** Is hazard prediction performance of a distracted driver affected by including visual or social cues showing positional information with a hazard alert?
- **Ch6 RQ3** Are distracted drivers better able to use attentional cues presented through a social agent to aid their hazard awareness compared to a visual cue?

6.3 Experiment 5

Comparison of Visual and Social Modalities for a Positional Cue on Hazard Prediction Ability while Distracted

Methods

Design

An experiment was conducted comparing situational awareness in drivers, as measured through hazard prediction ability, while engaged with an NDRT presented on an AR HUD. The experiment was designed to measure how hazard prediction ability was affected by the inclusion of a positional cue, either visual or social in nature, indicating the location of a hazard. A repeated measures experimental design was employed, with hazard prediction score, subjective confidence ratings and subjective attention ratings as dependent variables. The experiment used a one-way repeated measure design with 4 conditions: No NDRT (Control), NDRT only, Visual Cue, or Agent Cue. This was to measure whether drivers could use the implicit information provided from the positional cues while they are distracted by the NDRT to inform their choices on the hazard prediction task. The experimental design was approved through the College of Science and Engineering Research Ethics committee (Application number #300230057).

Participants

A power analysis was conducted using the *pwr* package (Champely et al., 2016) in R Studio, which determined that for a moderate effect size of 0.81, 21 participants were required for the repeated measures design. Twenty four participants *(Mean Age = 30.4 years, SD = 7.2, 11 Female)* were recruited via online forums and around the University of Glasgow Computer Science and Psychology departments. All had normal or corrected to normal eyesight and had held a driving licence for at least 2 years. Since previous research has shown the Hazard Prediction test to be culturally agnostic (Ventsislavova et al., 2019), this was not limited to drivers from the UK (10 UK, 4 Thailand, 4 India, 1 Bahrain, 1 China, 1 Germany, 1 Iran, 1 New Zealand, and 1 Spain). The average total driving experience was 10.41 years *(min = 3, max = 30, SD = 7.63)*, the average UK driving experience for non-UK license holders was 0.65 years *(min = 0, max = 6, SD = 1.63)*. Fourteen people reported they had experience of driving in the Greater Glasgow area where the hazard clips were filmed, with an average of 3.34 years *(min = 0, max = 20, SD = 5.51)*. Four reported having used an AR headset before, six reported using mobile AR and eleven reported never having used AR. Three participants reported never having heard of AR.

Figure 6.2: The experimental setup used for Experiment 5 and Experiment 6 with Hazard Prediction task displayed on the central monitor of the driving simulator and participants giving responses via buttons on the steering wheel and the mouse.

Hazard Prediction test

The same Hazard Prediction test (What Happens Next - WHN test) as presented in Chapter 4 was used to measure driver performance. Participants were presented with the 40 hazard clips whilst sitting in a driving simulator, with the test presented on the centre screen with the other two monitors turned off to avoid potential distraction. This test was built using PsychoPy v2021.2.3 (Peirce et al., 2019) and displayed on a Samsung LC32R500FHRXXU 32in curved monitor approximately 1 metre from participants. A multiple choice list of 4 potential scenarios were presented, which they selected using one of the shape buttons on a Logitech G29 steering wheel (see Figure 6.2).

Augmented Reality

The AR NDRT was developed in Unity (version 20203.26f1) using the Mixed Reality Toolkit (MRTK - version 2.7.2) and presented using the HoloLens 2 Augmented Reality headset (Microsoft, 2024). Guidelines for placing and displaying Mixed Reality content from Microsoft (2023) were followed for the placement, size and opacity. The same gem popping game as in Experiment 3 was used for the NDRT. Coloured gems would appear in 3D space in front of the computer monitor displaying the hazard prediction test. These gems appeared at random intervals between 1.5 and 2 seconds and participants were asked to 'pop' all the gems they could as quickly as possible by fixating on them.

They received a point for each gem popped, where performance was measured by counting the number of gems successfully popped against the overall number of gems spawned. The gems stopped spawning when the multiple choice options for the hazard prediction test appeared, allowing participants to focus on answering the question. This eye tracking task was confined to the area in front of the centre monitor displaying the video clips to emulate a HUD windscreen display.

Visual Cue

The Visual Cue condition was inspired by the LED display used by Trösterer et al. (2019). A blue bar object (approximately 7 cm wide) was displayed 1 metre away from participants at the bottom of the screen, which moved horizontally to highlight a particular area of the road. This was placed just below the main monitor in the video letterboxing but still in the drivers eyeline so the movement was visible as they were performing the NDRT. Four seconds prior to the hazard onset (based on findings from Eriksson and Stanton (2017)), the visual cue would move along the horizontal axis of the screen and highlight the area where the hazard was about to appear (see Figure 6.4).

Social Agent

The agent used for this experiment was selected through a short pilot study. Four potential virtual agents were shown observing the hazard prediction clips to 6 experts in the HCI and AutoUI domains. In Figure 6.3 from left to right: Agent 1 was chosen as a virtual recreation of the robot used by Karatas et al. (2015), Agent 2 was chosen to emulate a realistic human head, Agent 3 was chosen for its oblong shape which made observing the rotation easier, while Agent 4 was chosen to minimally intrude into the drivers view of the road (Bobor, 2024)[1].

After observing 5 clips of the agent viewing the road and responding to a hazard, participants were asked to rate each agent on the Robotic Social Attribute scale - (RoSAS) (Carpinella et al., 2017). Agent head 3, was rated as having the highest warmth and competence levels and the lowest discomfort levels after comparing each element of the RoSAS subscales, and so was selected for the main experiment. Like the visual cue, the virtual agent was placed just below the main monitor screen in the letterboxing of the video, so the rotation of the head was visible in the field of view of the headset as they performed the hazard prediction task but without obscuring the road (see Figure 6.4).

[1] Agent 1 was created in Unity to emulate the NAMIDA robot used by Karatas et al. (2015). Both Agent 2 and 3 were taken from the 3D library of Microsoft Paint 3D Design and are used with permission from Microsoft https://apps.microsoft.com/detail/9nblggh5fv99?hl=en-us&gl=US. Agent 4 was used under the Standard Unity Asset Store EULA https://assetstore.unity.com/packages/3d/characters/robots/sci-fi-drones-90326 (Bobor, 2024).

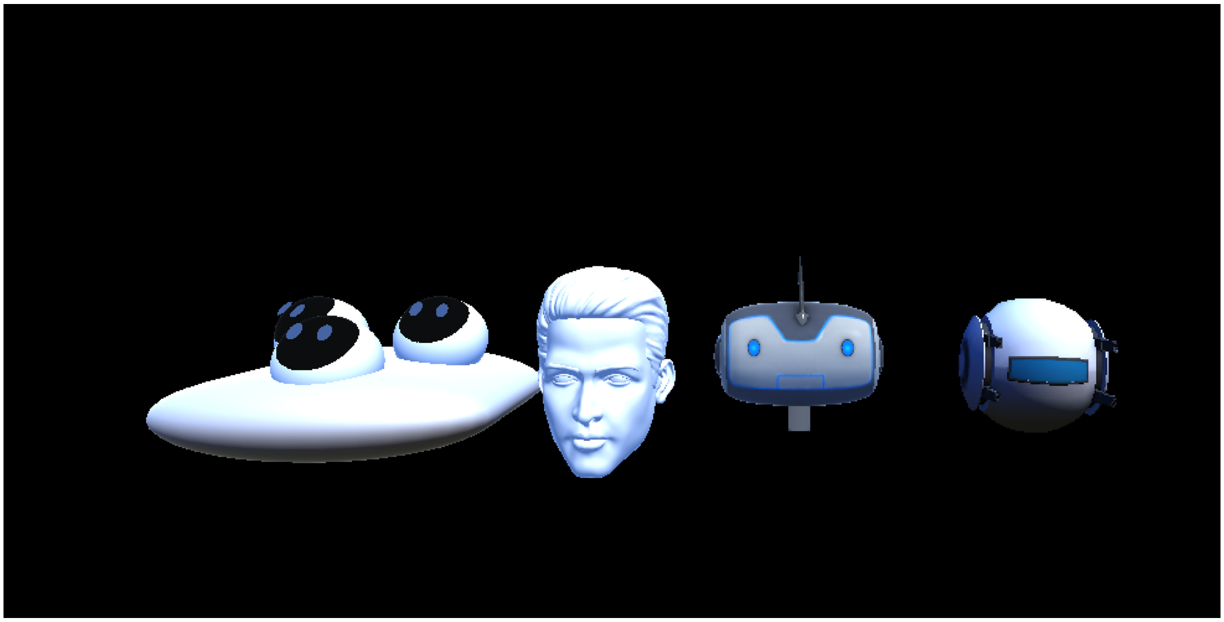


Figure 6.3: The four agents used in the pilot study from left to right: 1) three robot head based off Karatas et al. (2015), 2) a realistic human head 3) an agent with a rectangular head and 4) an agent with a round head.

Figure 6.4: Examples of the Experiment 6 conditions, with a) the Control (top left), b) NDRT only (top right), c) Visual that moved underneath the hazard (bottom left) and c) Agent which rotated towards the hazard (bottom right) conditions.

Both positional cues would periodically move to focus on non-hazardous objects in the road scene for 1500 ms as the hazard clips played, to familiarise participants with their function. Four seconds before the clip occlusion (based on average TOR reaction time findings from Eriksson and Stanton (2017)), both attentional cues would focus on the specific area where the hazard occurred, i.e., if the hazard was 'pedestrians step out from the left', the visual cue would move to the left side of the screen, while the virtual agent would rotate to look towards the left side of the screen (see Figure 6.4).

Procedure

Participants first filled in a demographics questionnaire as well as the Driver Behaviour Questionnaire (Reason et al., 1990). They then sat in the driving simulator and were shown an example of a Hazard Prediction clip to familiarise themselves with the main task. Following this, they were given the HoloLens 2 AR headset which was calibrated using the in-built calibration procedure. Participants proceeded through the four conditions (Control, NDRT only, Agent Cue, Visual Cue) in a counterbalanced order, with 24 total iterations. Before each condition, (except Control) participants were given a brief introduction to the AR components. For the NDRT conditions, this involved practising popping the gems whilst watching 2 example clips, whereas for the AR conditions this also involved observing the function of each attentional cue on the two practice clips. These were both described as *"demonstrating where the vehicle suggests you should be paying attention to"*. Each condition was presented as a block of 10 Hazard clips, totalling 40 clips for all conditions. Participants watched each clip and gave a response via a button on the steering wheel for each WHN clip, after which they were asked to rate their confidence in their responses on a 0-100 scale using a mouse. At the end of each condition, they rated their subjective attention to the driving task on a 0-100 scale, and filled in the NASA-TLX measuring their perceived workload for the completed condition. In the AR conditions, participants also completed the RoSAS questionnaire (Carpinella et al., 2017) rating how social they perceived the attentional cue (see Figure 6.5). At the end of the experiment, participants ranked each of the 4 conditions in order of preference and took part in a short semi-structured interview.

The study took around 60 minutes to complete and participants received a £10 voucher as compensation for taking part. In order to incentivise attention to both tasks, participants were told they could win an extra £5 voucher if they popped the most gems and answered the most WHN questions correctly compared to their peers.

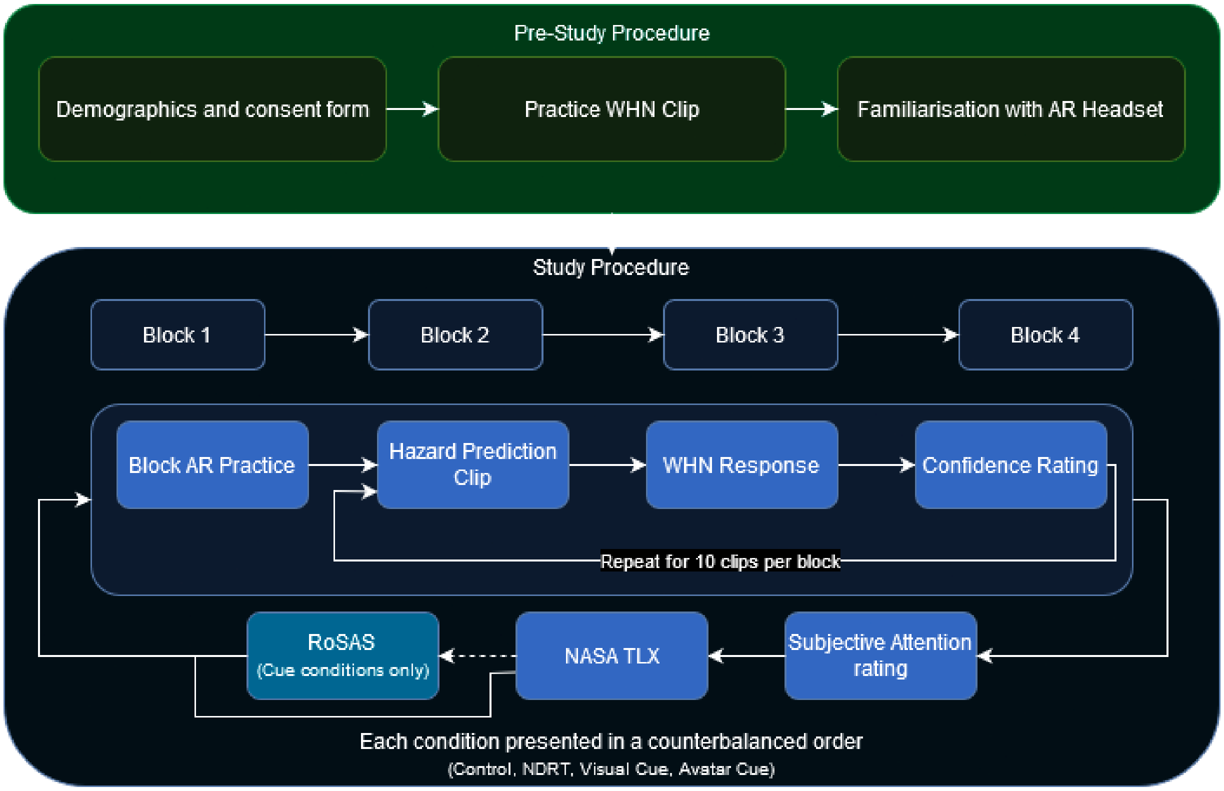


Figure 6.5: Diagram showing the procedure of the experiment. Each condition was presented in a counterbalanced order for each participant, with 10 Hazard Prediction clips per block.

Eye tracking

Eye movement data were collected using the Augmented Reality Eye Tracking Toolit - (ARETT) package for Unity 3D (Kapp et al., 2021) using the Hololens 2. A Screen object was placed in the scene made up of 9 Area of Interest - (AOI) components hidden from view, which split the screen into 9 distinct areas (see Figure 6.6). Participants were asked to calibrate the location of the virtual screen in the headset by adjusting the layout of the AR scene. This was done by moving a small floating cube which acted as a parent to the Screen AOIs, and lining up two virtual pillars with the real computer monitor. This was confirmed by the experimenter before the trials commenced (see Figure 6.6). All AR elements were labelled as AOIs using the 'Eyetracking' layer in ARETT. An AOI hit was measured when the Eyetracking object collided with an object in the Unity scene, with the location of the object, its name and the duration the objects were colliding was recorded (see Kapp et al. (2021) for a detailed description of the approach used).

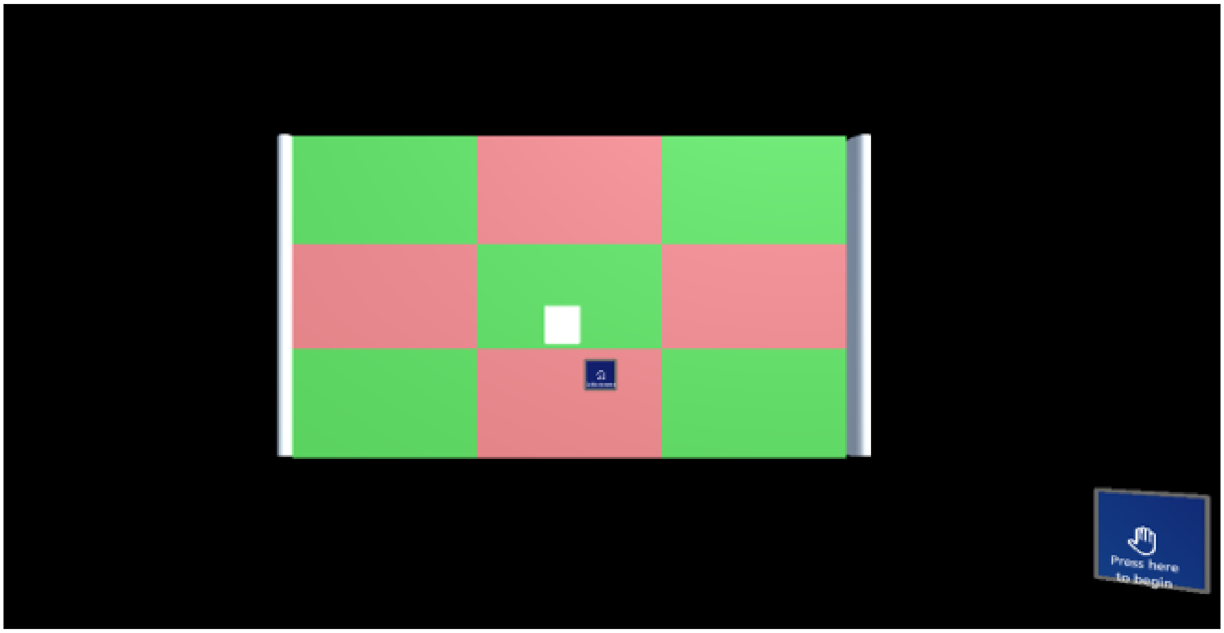


Figure 6.6: The 3x3 AOI grid object in Unity used to classify the location of fixations on the real life computer monitor. Collisions between and eyetracking object an AOI were classified as a hit.

6.4 Experiment 5 Results

The results were analysed using R Studio (2022.07.01 Build 554) using the lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017) and report (Makowski et al., 2023) R packages. Given the hierarchical nature of the data from the repeated measures design, Generalised Linear Mixed Effects - (GLME) models were fitted to the data for Hazard Prediction score and Confidence ratings, estimated using Maximum Likelihood - (ML) and the bobyqa optimizer. These models were determined through a backward stepwise model selection approach, where effects from more complex models that do not significantly explain a greater proportion of the variance are removed until arriving at the simplest model (see section 3.3). Confidence Intervals - (CI) of 95% and p-values were computed using a Wald t-distribution approximation. Repeated measures ANOVAs were conducted on the Attention and NDRT Performance scores as these were not nested data and so a mixed-effects models were not suitable.

Hazard Prediction

Average scores for the Hazard Prediction task for each of the 4 conditions (Control, NDRT, Agent & Visual) were compared (see Table 6.1). A GLME model was fitted to predict the main effects of Condition on Hazard Prediction score with a random intercept for each participant and a random intercept for each Hazard Clip, with the formula:

$$\text{Hazard Prediction Score}_{ij} = \beta._0 + \beta._1 \cdot \text{Condition}_{ij} + u_{0i} + u_{1j} + e_{ij}{}^2 :$$

The model's total explanatory power was moderate (conditional $R^2 = 0.33$). This model was found to explain significantly greater variance than a null model with Condition, Hazard clip and participant fitted as random effects *(Null Model AIC = 1089.6, BIC = 1104.2; Mixed Effects Model AIC = 1064.3,BIC = 1093.5;p < .001)*. Within this model, the score for the NDRT only *(Est. = -1.07, 95% CI [-1.54, -0.60], p < .001)*, Agent *(Est. = -1.19, 95% CI [-1.66, -0.72], p < .001*, and Visual *(Est. = -0.94, 95% CI [-1.41, -0.47], p < .001)* conditions were significantly lower than scores in the Control condition (see Figure 6.7). Refactoring the model with NDRT, Agent or Visual as intercepts produced no significant differences not already accounted for in the model described above (see Table 6.2 for a full list of model comparisons).

Block	Score (P)	Std Err	Lower CI	Upper CI
Control	0.87	0.03	0.79	0.92
NDRT only	0.69	0.05	0.57	0.78
Agent	0.66	0.06	0.54	0.763
Visual	0.71	0.05	0.60	0.80

Table 6.1: Summary statistics for average probability of a correct Hazard Prediction scores for each condition in Experiment 5

	Comparison											
WHN	Control			NDRT			Agent			Visual		
	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Control	1.86	(SE = 0.27)		-1.07	(SE = 0.24)	p <0.001***	-1.19	(SE = 0.24)	p = .004	-0.94	(SE = 0.24)	p <0.001***
NDRT				0.79	(SE = 0.25)		-0.12	(SE = 0.22)	p = .57	0.12	(SE = 0.22)	0.57
Agent							0.67	(SE = 0.25)		0.24	(SE = 0.22)	p = .26
Visual										0.92	(SE = 0.25)	

Table 6.2: Model estimates for Hazard Prediction Scores for Experiment 5, with the Standard Error - (SE) and p values for each of the generalised linear mixed effects models for each of the 4 conditions. Each row corresponds to a model with the named condition as the intercept, the column representing each of the other conditions compared to the intercept.

[2]- Hazard Prediction Score$_{ij}$ is the response variable for the i-th observation for the j-th participant. - $\beta._0$ is the fixed intercept. - $\beta._1$ is the fixed effect coefficient for the Condition variable. - Condition$_{ij}$ is the value of the Condition variable for the i-th observation for the j-th participant. - u_{0i} is the random intercept for the i-th participant, drawn from a normal distribution with mean zero and some participant-specific variance. - u_{1j} is the random intercept for the j-th hazard clip, drawn from a normal distribution with mean zero and some hazard clip-specific variance. - e_{ij} represents the residual error term for the i-th observation for the j-th participant. The fixed effects are denoted by $\beta.$ coefficients, and the random effects are represented by u terms.

Confidence Ratings

Average scores for Confidence ratings were compared for each of the conditions; Control, NDRT, Agent & Visual) (see Table 6.3). A GLME was fitted to predict the main effects of Condition with a random intercept for each participant and a random intercept for each Hazard Clip, with the formula:

$$\text{Confidence Rating}_{ij} = \beta_{\cdot 0} + \beta_{\cdot 1} \cdot \text{Condition}_{ij} + u_{0i} + u_{1j} + e_{ij}{}^{3} :$$

The model's total explanatory power was moderate (conditional $R^2 = 0.17$). This model was found to explain significantly greater variance than a null model with Condition, Hazard clip and participant fitted as random effects *(Null Model AIC = 1043.6, BIC = 1058.2; Mixed Effects Model AIC = 1031.3,BIC = 1060.5;p < .001)*. Within this model, the score for the NDRT only *(Est. = -0.89, 95% CI [-1.37, -0.41], p < .001)*, Agent *(Est. = -0.74, 95% CI [-1.22, -0.25], p = .003)*, Visual *(Est. = -0.89, 95% CI [-1.37, -0.41], p < .001)* conditions were significantly lower than in the control condition. Refactoring the model with the NDRT, Agent or Visual conditions as intercepts produced no significant differences not already accounted for in the model described above (see Table 6.4 for a full list of model comparisons).

Block	Rating	Std Err	Lower CI	Upper CI
Control	0.87	0.03	0.82	0.91
NDRT	0.74	0.04	0.65	0.81
Agent	0.76	0.04	0.68	0.83
Visual	0.73	0.04	0.65	0.81

Table 6.3: Summary statistics for average confidence ratings for each condition in Experiment 5, as well as the standard error and both lower and upper confidence intervals as reported from the mixed effects model.

	Comparison											
Confidence	Control			NDRT			Agent			Visual		
	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Control	1.91	(SE = 0.24)		-0.89	(SE = 0.25)	p <0.001***	-0.74	(SE = 0.25)	p = .003**	-0.89	(SE = 0.24)	p <0.001***
NDRT				1.02	(SE = 0.2)		0.15	(SE = 0.22)	p = .49	0.005	(SE = 0.25)	p = .98
Agent							1.17	(SE = 0.21)		-0.15	(SE = 0.22)	p = .17
Visual										1.02	(SE = 0.2)	

Table 6.4: Model estimates for Confidence Ratings for Experiment 5, with the Standard Error - (SE) and p values for each of the generalised linear mixed effects models for each of the 4 conditions.

[3]- Confidence Rating_{ij} is the response variable for the i-th observation for the j-th participant. - $\beta_{\cdot 0}$ is the fixed intercept. - $\beta_{\cdot 1}$ is the fixed effect coefficient for the Condition variable. - Condition_{ij} is the value of the Condition variable for the i-th observation for the j-th participant. - u_{0i} is the random intercept for the i-th participant. - u_{1j} is the random intercept for the j-th Hazard Clip - e_{ij} represents the residual error term for the i-th observation for the j-th participant. The fixed effects are denoted by β. coefficients, and the random effects are represented by u terms.

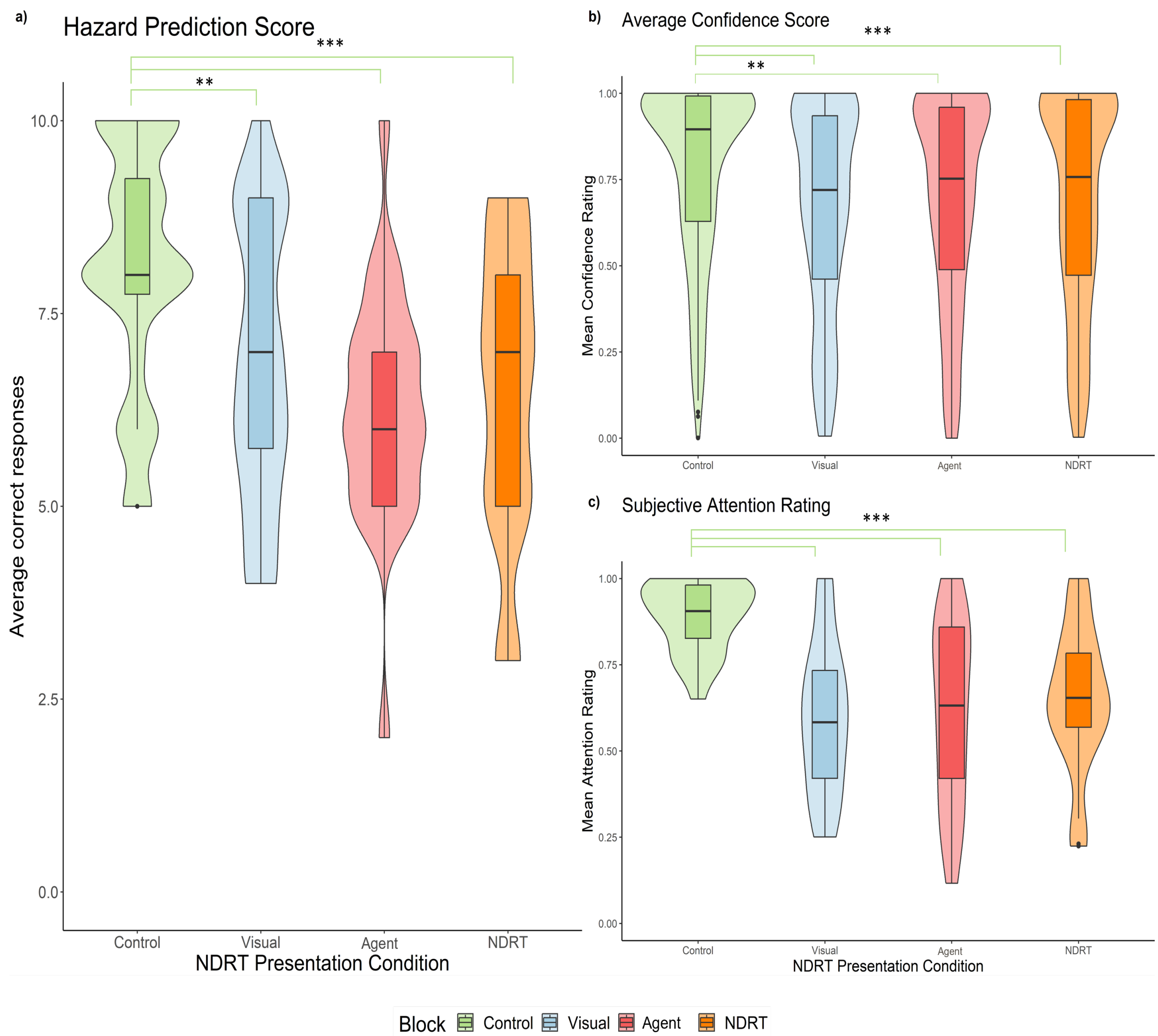


Figure 6.7: Average score on the Hazard Prediction task (left), average confidence ratings (top) and subjective attention ratings (bottom) for each condition in Experiment 5.

Subjective Attention Ratings

A one-way repeated measures ANOVA was conducted on the Subjective Attention ratings. As described above, a mixed effects model was deemed not suitable due to the nature of the data. The subjective attention rating was statistically significantly different for the different conditions *($F(3, 69) = 15.53$, $p < .001$, $\eta^2 = 0.26$)*. *Post hoc* analyses with a Bonferroni adjustment revealed that attention ratings in all AR conditions were significantly lower ($p < .001$) than Control in all conditions. However, none of the comparisons between the conditions were significantly different.

Perceived Workload

A one-way repeated measures ANOVA was conducted on the total NASA TLX ratings at each condition. The Total TLX rating was statistically significantly different across different blocks *($F(3, 69) = 10.498$, $p < .001$, $\eta^2 = 0.15$)*. *Post hoc* analyses with a Bonferroni adjustment revealed that the pairwise comparisons between the Control condition ratings were significantly lower than the NDRT *($p < .001$)*, Agent *($p = .02$)*, & Visual *($p < .001$)*. However, none of the comparisons between the AR conditions were significantly different (see Figure 6.8). For each of the 6 individual scales, there was the same pattern of results for the Mental Demand, Temporal Demand & Frustration subscales. For Overall Performance, there were only significant differences in the NDRT *($p = .004$)* and Visual *($p = .014$)* conditions, but not in the Agent condition. For Effort, there was only a significant difference in NDRT condition *($p = .016$)*. There were no significant differences with the Physical Demand subscale (see Appendix F for a full list of statistical comparisons).

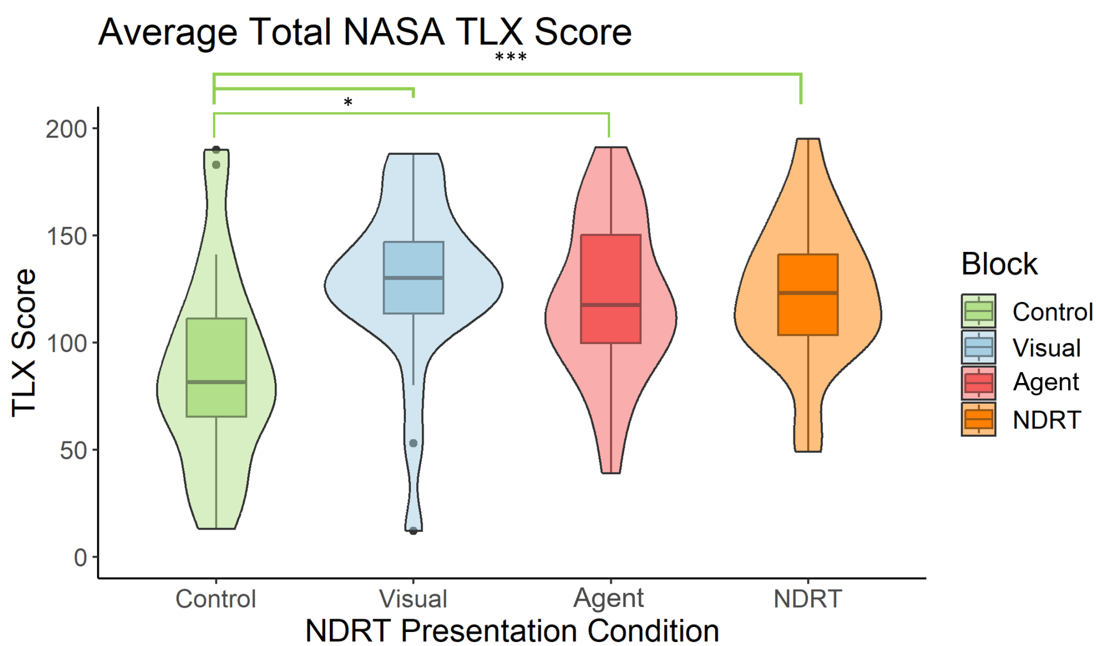


Figure 6.8: Average total NASA TLX ratings for each condition for Experiment 5.

NDRT Performance

A one-way repeated measures ANOVA was conducted on the task performance for the NDRT. However, there were no significant differences between average scores in any of the NDRT conditions.

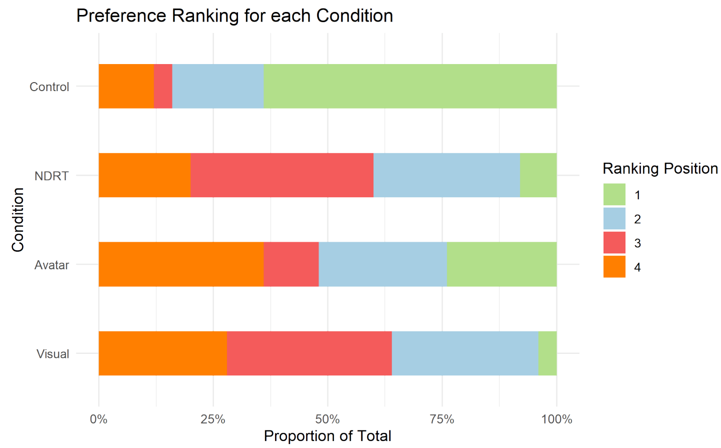


Figure 6.9: Preference ranking for each of the Conditions in Experiment 5.

Rankings

Participants were again asked to rank their preferred condition for completing the driving task in order of most to least favourite. These were converted into positional rankings for each user, where most preferred equalled to 4 points and least preferred equalled to 1 point. The rankings were then summed and weighted by the number of points each ranking received. The Control condition with no non-driving task was ranked the highest, followed by the Visual Cue condition, the Agent Cue condition with the Social Cue condition being the least preferred. See Table 6.5 for the full list of rankings and Figure 6.9.

Condition	1st	2nd	3rd	4th	Ranking
Control	64	15	2	3	**84**
Visual	4	24	18	7	53
Agent	24	21	6	9	60
NDRT only	8	24	20	5	57

Table 6.5: Participant rankings of each Cue condition with the number of points gained for each ranking in Experiment 5. Participants were asked to rank the conditions in order of preference for being able to complete both the driving and non-driving tasks.

RoSAS

A series of one-way repeated measures ANOVAs were conducted comparing ratings on the RoSAS scale between the Visual and Agent condition. For the Warmth factor, there were significant differences between ratings on the Happy $(F(1,23) = 9.24,\ p = .006,\ \eta^2 = 0.17)$, Social $(F(1,23) = 9.24,\ p = .006,\ \eta^2 = 0.17$, Compassionate $(F(1,23) = 4.86,\ p = .038,\ \eta^2 = 0.06)$ & Emotional subscales $(F(1,23) = 4.67,\ p = .041,\ \eta^2 = 0.07)$, where the Agent was rated higher than the Visual cue. For the Competence factor, there was a significant difference on the Responsive subscale $(F(1,23) = 4.92,\ p = .037,\ \eta^2 = 0.08)$ again with the Agent condition rated higher than the Visual cue (see Table F.2).

Eyetracking

All analysis was conducted using the ARETT package (Kapp et al., 2021). A total of 13,335,248 gaze points were recorded across all participants in all four conditions before processing. The I-AOI fixation classification function from the ARETT R package (Kapp et al., 2021) was used to classify valid fixations. This function, based on Salvucci and Goldberg (2000)'s definition, classifies relevant fixations based on whether they land on a pre-defined AOI layer in Unity for longer than a predetermined threshold. Fixations longer than 60ms that hit either the computer screen (recorded as one AOI in a 3x3 grid - See Figure 6.6) or any 3D object in AR were retained for analysis. Given the high level of variability between participants that comes from using a portable AR headset, this method was selected over more spatially accurate classification algorithms which are sensitive to head movements, which Kapp et al. (2021) and Aziz and Komogortsev (2022) recognise as a limitation of in spatial accuracy of the HoloLens 2. Two participants were removed due to technical issues with the data collection, leaving 538,429 valid fixations for analysis from 22 participants.

Total Fixation Count

A Linear Mixed Effects - (LME) model was fitted comparing the mean number of fixations for each condition, with participants as a random effect. The model had substantial explanatory power (conditional $R^2 = 0.61$). Overall, there were a significantly lower number of fixations in the Control condition compared to Visual *(Est. = 113.33, 95% CI [76.41, 150.26], t(78) = 6.11, p < .001* Agent *(Est. = 156.86, 95% CI [119.93, 193.78], t(78) = 8.46, p < .001)* and NDRT *(Est. = 126.00, 95% CI [89.07, 162.93], t(78) = 6.79, p < .001)*. There were also significantly fewer fixations in both the Visual condition *(Est. = -7.48, 95% CI [-13.68, -1.28], t(78) = -2.40, p = .019)* and NDRT conditions *(Est. = -8.33, 95% CI [-14.53, -2.13], t(78) = -2.68, p = .009)* when compared to the Agent condition. However, no other comparisons produced significantly different results (see Table 6.6 and Figure 6.10).

Hazard Fixation Count

A LME model was fitted comparing the mean number of fixations during the last 4 seconds of the hazard clips for each condition. The model had substantial explanatory power (conditional $R^2 = 0.52$). Overall, there were significantly fewer fixations in the Control condition compared to Visual *(Est. = 16.19, 95% CI [9.99, 22.39], t(78) = 5.20, p < .001)* Agent *(Est. = 23.67, 95% CI [17.47, 29.87], t(78) = 7.60, p < .001)* and NDRT *(Est. = 15.33, 95% CI [9.13, 21.53], t(78) = 4.92, p < .001)* conditions. No other comparison produced significantly different results (see Table 6.6 and Figure 6.10).

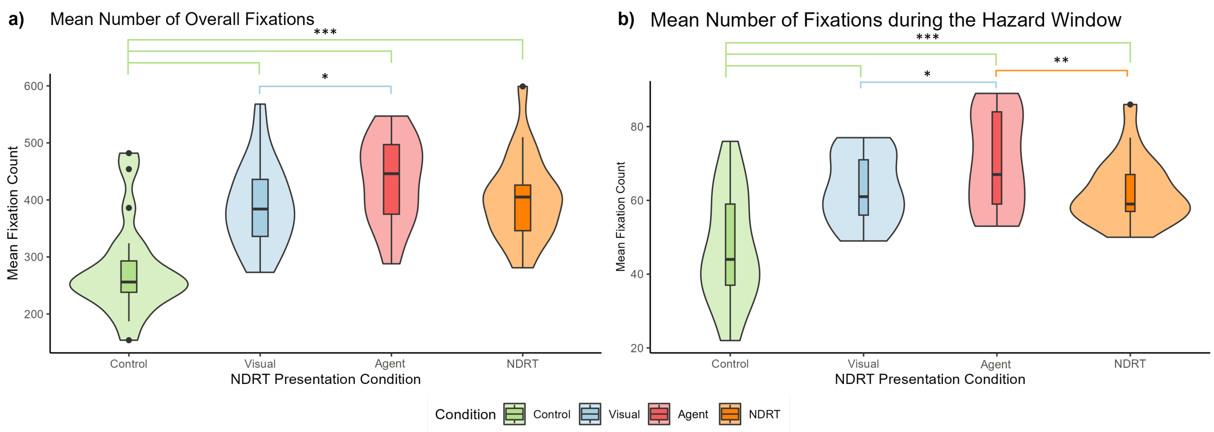


Figure 6.10: a) The average total number of fixations in each condition and b) the average number of fixations during the hazard window in each condition for Experiment 5. The correct area was determined as whether a fixation landed on the correct portion of the screen in the 9x9 AOI grid for at least 100ms during the last 4 seconds of each clip.

Condition	Total Fixation		Hazard Window		Correct Fixations		Proportion Correct	
	Average	SD	Average	SD	Average	SD	Percentage	SD
Control	276.95	80.64	46.9	14.94	22.52	6.65	50.23	12.24
NDRT	402.95	73.8	62.24	8.73	21.33	5.89	34.76	10.6
Agent	433.81	73.96	70.57	12.8	23.71	4.96	33.78	4.8
Visual	390.29	74.59	63.1	9.28	21.48	6.52	34.33	10.21

Table 6.6: Summary of the average total number of fixations in each condition overall, during the Hazard Window, the average number of correct fixations during the hazard window and the average proportion of correct fixations in each condition for Experiment 5.

Correct Hazard Fixations

'Correct' Fixations were classified depending on whether they were located in the same AOI on the screen as the hazard in the video clip, during the period up to four seconds before the end of the Hazard Prediction clip. This was the length of time that the cues were indicating the location of the hazard. A LME model was fitted to predict average number of fixations in Correct AOI based on Condition with participant as a random factor. However, there were no significant differences in the number of correct fixations during the Hazard window between any of the conditions (See Figure 6.11). There was a significantly higher overall proportion of Correct fixations in the Control condition compared to the Visual *(Est.* $= -15.90$*, 95% CI [-20.93, -10.86],* $t(78) = -6.28$*,* $p < .001$, Agent *(Est.*$= -16.46$*, 95% CI [-21.50, -11.42],* $t(78) = -6.50$*,* $p < .001$*)* and NDRT *(Est.* $= -15.47$*, 95% CI [-20.51, -10.43],* $t(78) = -6.11$*,* $p < .001$*)* conditions. However, there were no significantly different results from other comparisons (see Table 6.6 and Figure 6.11).

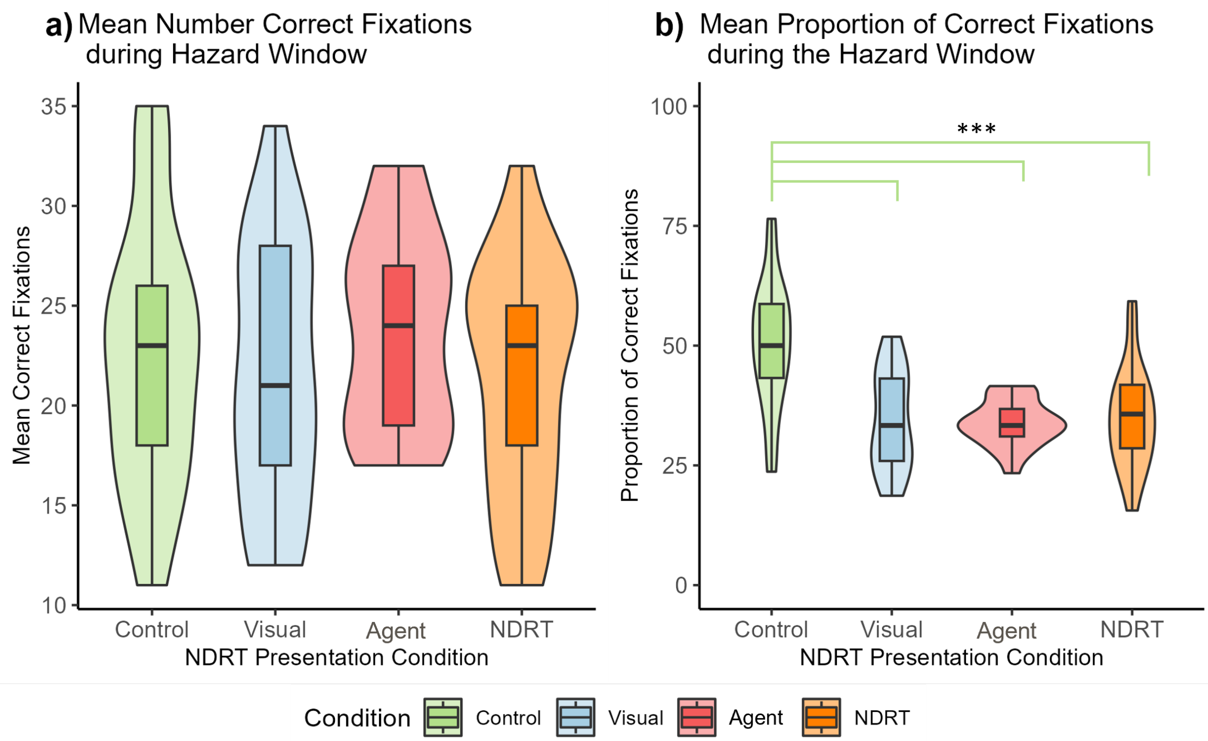


Figure 6.11: a) The mean number of correct fixations on the hazard and b) the mean proportion of correct fixations on the hazard score for Experiment 5.

Total Fixation Duration

Due to the non-normal distribution of the data, fixation duration was z-scored to check for outliers and 94,430 fixations which fell above 3 standard deviations from the mean duration were excluded from the analysis. Similarly, 3764 fixations which did not have a registered gaze point target or AOI target were also removed, since the lack of a relevant target meant that the participant was fixating on neither the screen nor any of the AR content.

A LME model was fitted to model the effect of Condition on average fixation duration with participant as a random effect that had substantial explanatory power (conditional R^2 = 0.81). In the Control condition, there were significantly longer overall fixation durations compared to the Visual *(Est. = -386.56, 95% CI [-459.30, -313.82], t(78) = -10.58, p < .001)*, Agent *(Est. = -502.54, 95% CI [-575.28, -429.80], t(78) = -13.75, p < .001)* and NDRT *(Est. = -393.77, 95% CI [-466.51, -321.03], t(78) = -10.78, p < .001)* conditions. Overall Fixation Duration was significantly shorter in the Agent condition compared to the Visual*(Est. = -115.98, 95% CI [43.24, 188.72], t(78) = 3.17, p = .002)* and NDRT *(Est. = 108.77, 95% CI [36.03, 181.51], t(78) = 2.98, p = .004)* conditions. No other comparison produced significantly different results (see Table 6.7 and Figure 6.12).

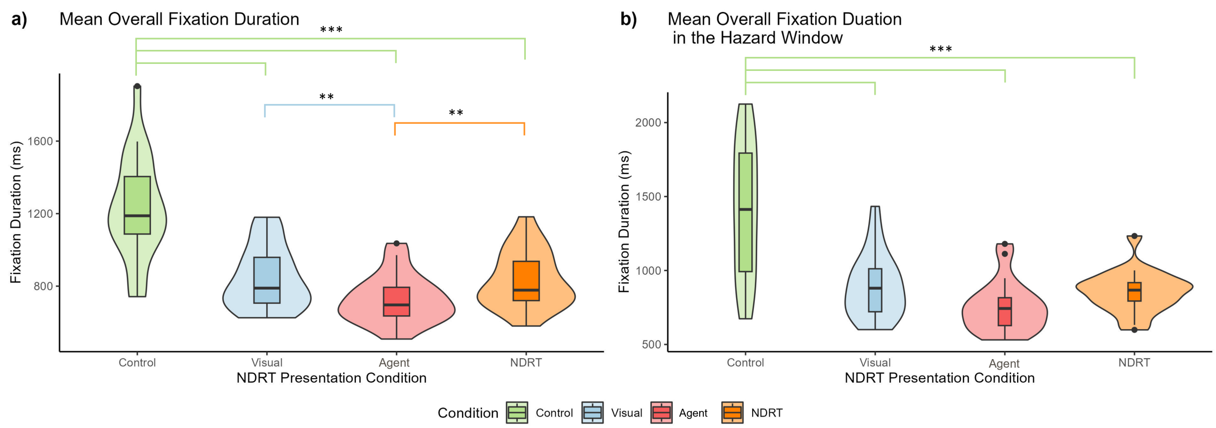


Figure 6.12: a) The mean fixation duration throughout the whole experiment for each condition and b) the mean fixation duration in the last 4 seconds before the hazard appeared score for Experiment 5.

Condition	Mean Fixation Duration (ms)		Mean Hazard Fixation Duration (ms)		TTFF (ms)	
	Mean	SD	Mean	SD	Mean	SD
Control	1165.37	1368.45	1408.67	170.92	2034.68	1616.84
NDRT	811.64	859.34	856.93	140	2185.02	1556.57
Agent	706.99	732.86	753.24	170.92	2361.64	1541.29
Visual	815.91	846.39	894.95	215.58	2280.14	1547.95

Table 6.7: A table showing the mean duration of fixations in each condition overall, mean durations during the Hazard Window and the Time to the First Fixation (TTFF) in Experiment 5.

Hazard Fixation Duration

As before, 'correct' fixations were classified depending on whether they hit the same AOI area of the screen as the hazard up to four seconds before the end of the Hazard Prediction clip. A LME model was fitted to predict fixation length based on Condition with participants as a random effect, which had substantial explanatory power (conditional R^2 = 0.57). Within this model, there were significantly longer fixation duration in the Control condition compared to the Visual *(Est. = -513.73, 95% CI [-663.00, -364.45], t(78) = -6.85, p <.001)*, Agent *(Est. = -655.43, 95% CI [-804.70, -506.15], t(78) = -8.74, p < .001)* and NDRT *(Est. = -551.75, 95% CI [-701.02, -402.47], t(78) = -7.36, p <.001)* conditions. However, there were no significant differences between any of the other conditions (see Figure 6.12).

Time to First Fixation

A LME model was fitted to model the effect of Condition on time until the first fixation on the correct area of the screen. There were significantly slower first fixations in the Visual *(Est. = 0.214, t(1871) = 2.222, p = .0264)* and Agent *(Est. = 0.2143, t(1901) = 2.069, p = .039)*, but no difference with the Social condition. However, there were no significant differences between any of the other conditions (see Table 6.15 and Figure 6.13).

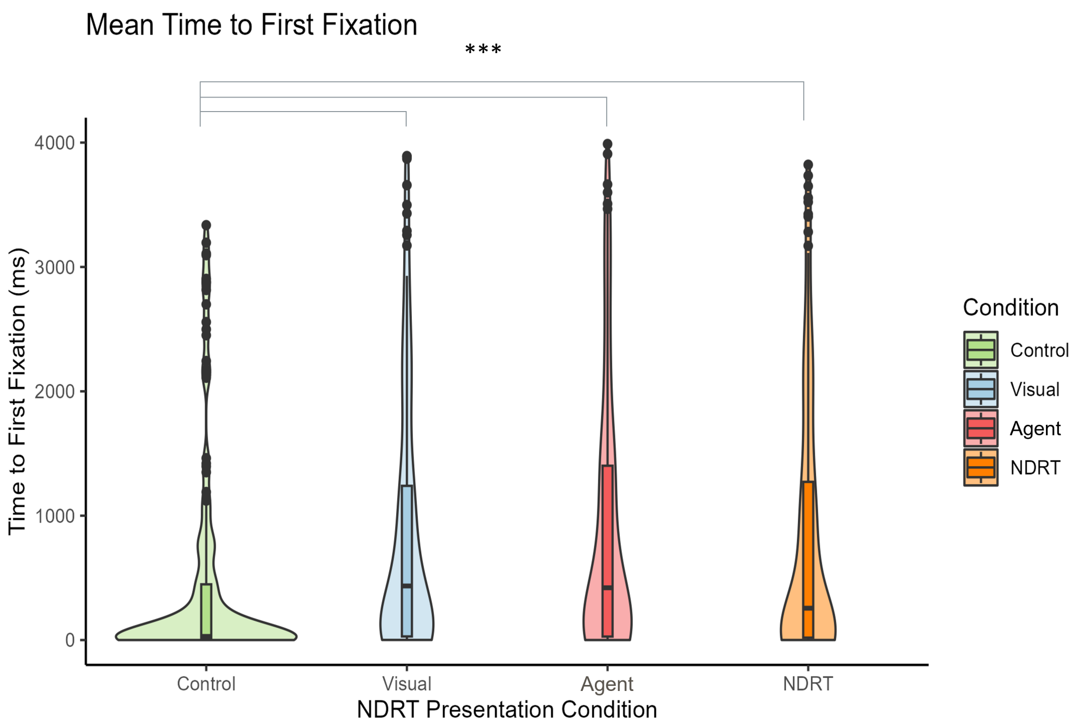


Figure 6.13: Average Time to First Fixation on the hazard for each condition in Experiment 5. The Control condition graph tending towards 0 indicates that the first fixations of participants in this group were before the 4 seconds Hazard window where the cues appeared in the other conditions.

6.5 Experiment 5 Discussion

In all conditions, hazard prediction scores were higher than chance, indicating that participants were able to maintain some situational awareness of the events in the clips. However, performance in the Control condition, where attention was solely on the road, was significantly higher than all other conditions with an NDRT. There were no differences between scores in any of the three NDRT conditions, regardless of the inclusion of an attentional cue. The same pattern was true for the confidence and subjective attention ratings, with higher ratings in the Control conditions than the conditions with an NDRT, regardless of the inclusion of the cues. This is despite the cues reliably highlighting the location of the hazard in each clip. These results suggest that, without a clear signal to indicate that a dangerous event is about to occur, the positional cues provided no additional benefit to hazard prediction ability. It is well established that distraction can lead to missing seemingly obvious information (Simons and Chabris, 1999; Drew et al., 2013; Gelderblom and Menge, 2018), which is the pattern observed in these results. Without any attention capturing aspect, these 'nudging' cues were not effective in aiding awareness of the road, resulting in increased the perceived workload and impairing hazard prediction ability when drivers were distracted.

The NASA TLX data support this, as there were no differences in perceived workload or confidence ratings between the cue conditions and the NDRT only conditions. Similarly, the eye tracking data indicate that distracted drivers were looking at the correct area of the screen at the same rate as when viewing the clips without an NDRT. In fact, the time for drivers to first fixate on the correct area of the screen was delayed in all NDRT conditions compared to control, indicating that the distracting nature of the NDRT directly impacted the ability to switch to the driving task. Like in Experiment 4 however, this did not correspond to increased hazard prediction performance. Overall, the eye tracking data indicate that the NDRT was disruptive to the normal gaze behaviour when viewing the hazard prediction clips as in the Control condition. There were fewer fixations in the NDRT conditions overall and during the hazard window, with shorter fixation durations when compared to the Control condition, which is indicative that drivers were sufficiently processing the scene (Chapman and Underwood, 1998; Crundall and Underwood, 2011).

The results from this experiment suggest that the agent's head turning behaviour by itself, which has been used to cue attention in human-robot interactions (Yamazaki et al., 2008; Sheikhi and Odobez, 2015; Hoque et al., 2011), is not effective for cueing location when a driver is distracted by an NDRT. However, the same was true for the Visual condition, indicating that this is likely due to the lack of attention capturing aspect to the cue, when drivers were focused on performing the NDRT in the HUD. Without a specific cue to signal when to look at the road, the presence of the agent made no difference to hazard prediction scores compared to the NDRT only condition. This is an unlikely application for this type of interface though, as most attentional alerts and TORs include a specific *attention-capturing* aspect to alert the driver that there is danger (Karatas et al., 2020; Rusch et al., 2013). These types of alerts were not present in the positional cues used here, and thus there were no perceivable difference between the positional cues highlighting non-hazardous and hazardous areas of the road. This poses the question of whether these cues would have any benefit to distracted drivers if an additional hazard alert cue to signal danger is included? Can drivers use information from a social cue to aid their hazard awareness if paired with an explicit hazard warning in the same way as a visual cue? Additionally, while RoSAS ratings for the agent were higher on the Warmth sub-scales, there were no significant differences for the Competence sub-scale, other than it being seen as more "Responsive". Designing realistic social cues is key for establishing a social interaction (Admoni and Scassellati, 2017; Ruhland et al., 2015) and realistic interactions with robots increase positive perceptions of usefulness (De Wit et al., 2023; Groechel et al., 2019). Thus, it is also important to investigate whether a more socially expressive cue aids hazard prediction further or affects perceptions of the agent.

6.6 Experiment 6

Comparison of Visual and Social Modalities for an Active Attentional Cue on Hazard Prediction Ability while Distracted

A follow-up study was conducted to evaluate the effect of including attention capturing elements that signal danger into the positional cues used in Experiment 5. A visual cue that used position and colour change to indicate the location of the hazard, an agent that turned its head towards the danger and changed colour, and a social agent with an animated social reaction that gestured towards the danger, were compared to a Control condition where participants only focused on the hazard prediction task.

Methods

Design

The same design as Experiment 5 was employed to measure how Hazard Prediction ability was affected by inclusion of a positional attention cue with a hazard alert, either visual or social in nature. A repeated measures experimental design was used, with Hazard Prediction score and subjective confidence rating as dependent variables, as well as subjective attention, RoSAS, and NASA TLX scores. The independent variable was the attentional cue, with levels: no NDRT (Control), a colour changing visual hazard cue (*Visual Hazard Cue*), an agent with a colour-change hazard cue (*Agent Hazard Cue)*, or an agent exhibiting an animated social cue (*Social Hazard Cue).* The experiment took around 60 minutes to complete and was approved by the College of Science and Engineering Research Ethics Committee (#300230120)

Participants

A power analysis was conducted using the *pwr* package (Champely et al., 2016) in R Studio, which determined that for a moderate effect size of 0.81, 21 participants were required for the repeated measures design. Twenty four new participants *(Mean Age = 29.75 years, SD = 10.78, 12 Female)* were recruited via online forums and around the University of Glasgow Computer Science and Psychology departments. All had normal or corrected to normal eyesight and had held a driving licence for at least 2 years. Again, this was not limited to drivers from the UK (13 UK, 2 China,1 Australia, 1 Austria, 1 Canada, 1 Cyprus, 1 Germany, 1 Italy, 1 Kenya, 1 Pakistan and 1 Slovenia). The average total driving experience was 10.34 years *(min = 2, max = 44, SD = 10.38)*, the average UK driving experience for non-UK license holders was 0.5 years *(min = 0, max = 4.16, SD = 1.23)*.

Eleven people reported they had experience of driving in the Greater Glasgow area where the hazard clips were filmed, with an average of 2.54 years *(min = 0.24, max = 11, SD = 3)*. Eight reported having used an AR headset before, five reported using mobile AR and ten reported never having used AR. One participant reported never having heard of AR.

Augmented Reality

The same AR NDRT was used as in Experiment 5. For the Visual Hazard Cue condition, the same blue bar from the Visual Cue condition in Experiment 5 was used, but changed colour from blue to red (taking between 1 and 2 frames to transition), 4 seconds before the occlusion of the clip. For the Agent Hazard Cue and Social Hazard Cue conditions, the agent design was based on the one from Experiment 5. In the Agent Hazard Cue condition, this agent was augmented so that the sides of the head and the antenna turned red 4 seconds before the end of clip. This design was taken from previous work which established colour change as a key attention capturing element (Hollingworth and Hwang, 2013; Wang et al., 2023), and the association of the colour red with danger (Pravossoudovitch et al., 2014; Chapanis, 1994), as well as designs from the focus group in Chapter 5. For the Social Hazard Cue condition, the agent was animated to draw attention to an area of the screen with a social cue. This animation was decided through a pilot questionnaire (N = 18), where 5 prototype social reactions were shown to participants based on different animated emotions (see Figure 6.14. Pilot participants were asked to write *"What emotion(s) do you see represented here?"* in an open text format. These responses were summarised into themes which described the common descriptions (see Figure 6.14 for these descriptions and examples). The final animation was selected for its consistency of responses and appropriateness in the pilot survey. It involved the agent shaking left and right to mimic quivering, with sweat droplets being emitted from the top of the head to suggest fear or apprehension. The agent also was given virtual hands that were raised to point towards the direction that the head was turning, to contextualise the cue as a social response to the hazard (De Wit et al., 2023). These hands were also present in the Agent Hazard Cue condition for consistency, but did not move or serve any cueing function.

Procedure

The same procedure as Experiment 5 was used (see Figure 6.5). Once again before each condition except the Control, participants were given a brief introduction to the AR components which involved observing the function of each cue on two practice clips highlighting the presence of a hazard. These were both described as *"the vehicle highlighting a dangerous element in the scene for you"*.

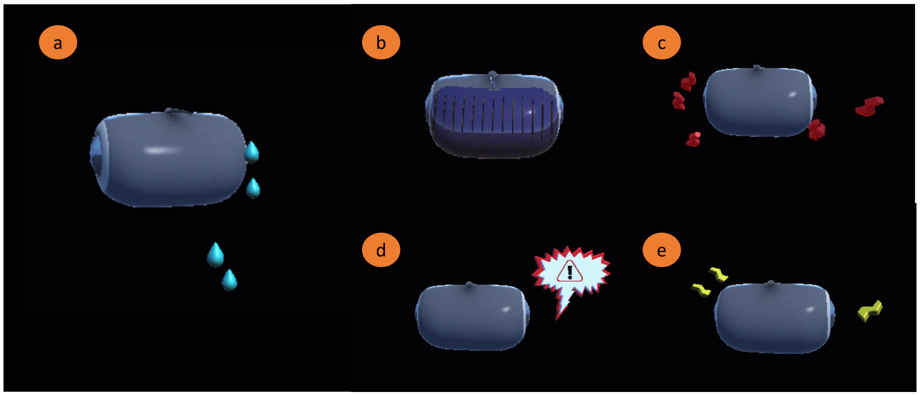


Figure 6.14: Images of the prototype animations for the social cue and their summary descriptions, with a) Sadness, Stress and Nervousness - sweat droplets falling down, b) Anger and Confusion - Fear and Excitement - black lines on blue shading, c) red lightning bolts that jittered, d) Danger and Urgency - a speech bubble with a warning symbol and e) Confusion and Frustration - yellow lightning bolts that were emitted in a wave pattern.

Eye tracking

Eye movement data were again collected using the ARETT package for Unity 3D (Kapp et al., 2021) using the HoloLens 2, following the same procedure as Experiment 5.

6.7 Experiment 6 Results

The same analysis procedure as Experiment 5 (see section 6.4) was followed.

Hazard Prediction

Average scores for the Hazard Prediction task for each of the 4 conditions (Control, Visual Hazard Cue, Agent Hazard Cue & Social Hazard Cue) were compared (see Table 6.8). A LME model was fitted to predict the main effects of Condition on Hazard Prediction score with a random intercept for each participant and a random intercept for each Hazard Clip, with the same formula and model selection approach as in Experiment 5 (see section 6.4).

The model's total explanatory power was substantial (conditional $R^2 = 0.34$). This model was found to explain significantly greater variance than a null model with Condition, Hazard clip and participant fitted as random effects *(Null Model AIC = 1020.6, BIC = 1034.7; Mixed Effects Model AIC = 1014.2,BIC = 1043.4;p = .007)*. Within this model, there were no significant differences between Hazard Prediction performance in the Control condition and the Visual Hazard Cue nor Agent Hazard Cue conditions. However, scores in the Social Hazard Cue condition were significantly lower than the Control *(Est. = 0.80, 95% CI [0.34, 1.26], p < .001)*, Agent Hazard Cue *(Est. = 0.46, 95% CI [0.02, 0.91], p = .040)* and Visual Hazard Cue *(Est. = 0.49, 95% CI [0.05, 0.94], p = .029)* conditions (see Table 6.8).

Figure 6.15: Examples of the Experiment 6 conditions, with a) the Control (top left), b) Visual Cue that turned red and moved underneath the hazard (top right), c) Agent Cue which turned red will looking towards the hazard (bottom left) and D) Social Cue which gesticulated and pointed towards the hazard.

Refactoring the model with NDRT, Agent or Visual as intercepts produced no significant differences not already accounted for in the model described above (see Figure 6.16 and Table 6.9 for a full list of model comparisons) [4]

Block	Score (P)	Std Err	Lower CI	Upper CI
Control	0.84	0.04	0.75	0.9
Agent Hazard Cue	0.78	0.05	0.68	0.86
Social Hazard Cue	0.7	0.06	0.58	0.79
Visual Hazard Cue	0.79	0.05	0.69	0.86

Table 6.8: Summary statistics for average percentage of correct Hazard Prediction scores for each Presentation method in Experiment 6, as well as the standard error and both lower and upper confidence intervals as reported from the mixed effects model.

[4]Given the pattern of results and the presence of an outlier in the Control condition, the model was also run with this outlier participant's data removed. However, since the same pattern of significance was observed and the outlier was not above 3 standard deviations away from the mean, it was decided to include the full dataset in the final model presented here.

	Comparison											
WHN	Control			Visual Hazard Cue			Agent Hazard Cue			Social Hazard Cue		
	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Control	1.627	(SE = 0.28)		-0.31	(SE = 0.24)	p = .2	-0.34	(SE = 0.24)	p = .164	-0.8	(SE = 0.23)	p <0.001***
Visual Hazard Cue				1.32	(SE = 0.27)		-0.029	(SE = 0.23)	p = .9	-0.49	(SE = 0.23)	p = .029*
Agent Hazard Cue							1.29	(SE = 0.27)		-0.464	(SE = 0.24)	p = .04*
Social Hazard Cue										0.826	(SE = 0.27)	

Table 6.9: Model estimates for Hazard Prediction Scores for Experiment 6, with the Standard Error (SE) and p values obtained through Wald's approximation for each of the generalised linear mixed effects models for each of the 4 cue conditions. Each row corresponds to a model with the named cue condition as the intercept, the column representing each of the other cue conditions compared to the intercept. Repeat comparisons were omitted for clarity, but represent the inverse of the estimate presented.

Confidence Ratings

Average scores for confidence ratings in Experiment 6 were compared for each of the NDRT cue conditions (Control, Visual Hazard Cue, Agent Hazard Cue & Social Hazard Cue) (see section Table 6.10). A linear mixed model was fitted to predict the main effects of Condition with a random intercept for each participant and a random intercept for each Hazard Clip, with the same formula as Experiment 5 (see 6.4). However, this model was not found to explain a significantly greater amount of variance than a null model with Condition, Hazard clip and participant fitted as random effects and so was rejected. See Table 6.11 for a full list of model comparisons.

Block	Ratings	Std Err	Lower CI	Upper CI
Control	0.78	0.03	0.78	0.83
Agent Hazard Cue	0.78	0.03	0.66	0.78
Social Hazard Cue	0.74	0.03	0.68	0.78
Visual Hazard Cue	0.75	0.03	0.69	0.8

Table 6.10: Summary statistics for average confidence ratings for each Presentation method in Experiment 6, as well as the standard error and both lower and upper confidence intervals as reported from the mixed effects model.

	Comparison											
Confidence	Control			Visual Hazard Cue			Agent Hazard Cue			Social Hazard Cue		
	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.	Estimate	SE	Sig.
Control	1.69	(SE = 0.21)		0.12	(SE = 0.26)	p = .643	-0.39	(SE = 0.24)	p = .1	-0.15	(SE = 0.25)	p = .53
Visual Hazard Cue				1.8	(SE = 0.22)		-0.51	(SE = 0.24)	**p = .04***	-0.27	(SE = 0.25)	p = .27
Agent Hazard Cue							1.29	(SE = 0.19)		0.24	(SE = 0.23)	p = .31
Social Hazard Cue										1.53	(SE = 0.2)	

Table 6.11: Model estimates for Confidence Ratings for Experiment 6, with the Standard Error (SE) and p values obtained through Wald's approximation for each of the generalised linear mixed effects models for each of the 4 cue conditions. Each row corresponds to a model with the named cue condition as the intercept, the column representing each of the other cue conditions compared to the intercept. Repeat comparisons were omitted for clarity, but represent the inverse of the estimate presented.

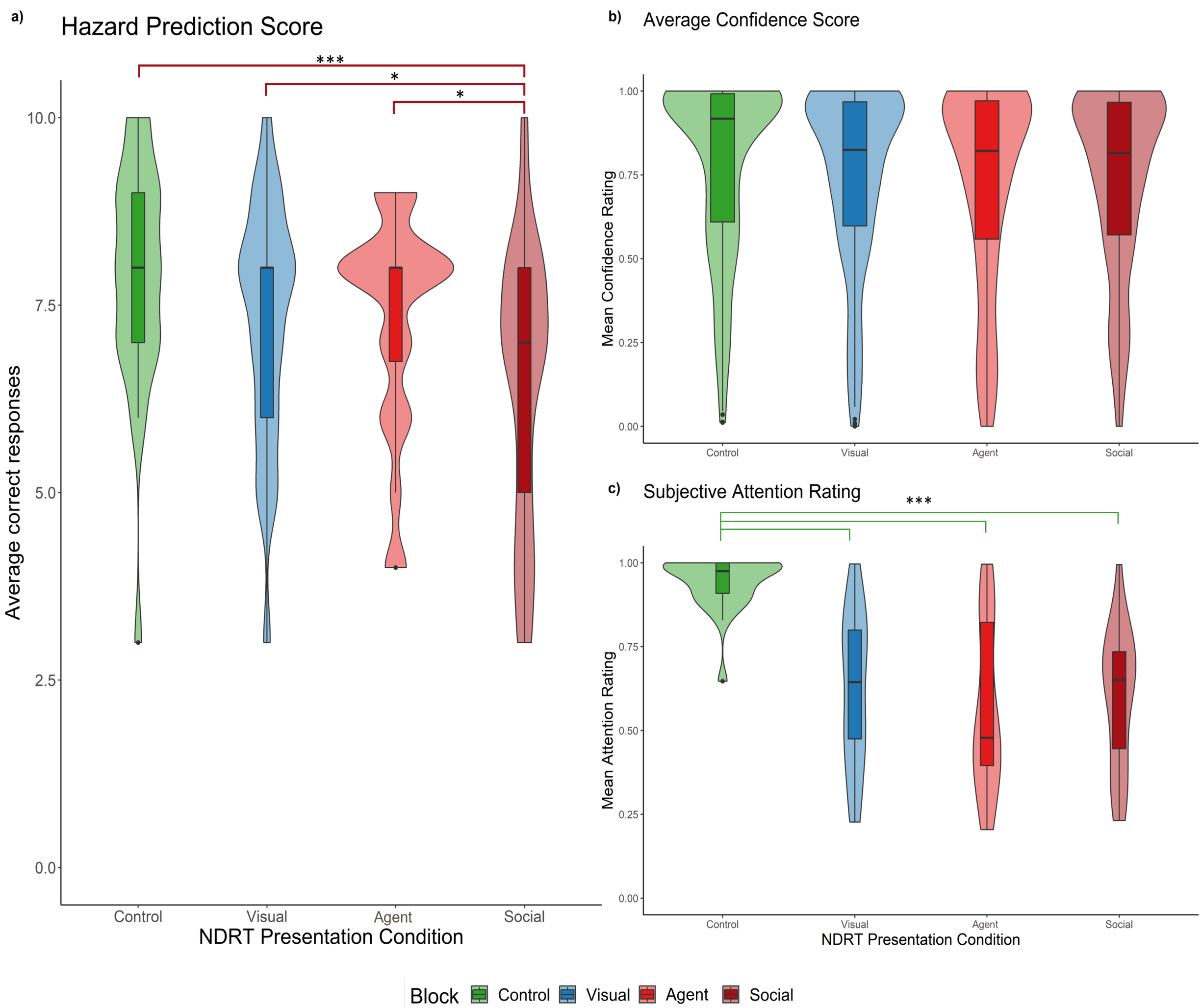


Figure 6.16: Average score on the Hazard Prediction task (left) average confidence ratings (top) and average subjective attention ratings (bottom) for each cue condition in Experiment 6.

Subjective Attention Ratings

A one-way repeated measures ANOVA was conducted on the subjective attention ratings, with participants as a random factor. These ratings were statistically significantly different at the different time points *($F(3, 69) = 37.44$, $p < .001$, $\eta^2 = 0.38$)*. *Post hoc* analyses with a Bonferroni adjustment revealed that attention ratings in all AR conditions were significantly lower *($p < .001$)* than Control in all conditions, mirroring the results from Experiment 5. Once again however, none of the comparisons between the AR conditions were significantly different.

Perceived Workload

A one-way repeated measures ANOVA was conducted on the overall NASA TLX ratings at each condition. The overall TLX rating was statistically significantly different across different conditions *($F(3, 69) = 4.6$, $p = .005$, $\eta^2 = 0.06$)*. *Post hoc* analyses with a Bonferroni adjustment revealed that the Control condition ratings were significantly lower than the Agent *($p = .02$)* condition. However, none of the other comparisons between the AR conditions were significantly different (see Figure 6.17). For each of the 6 individual scales, there was the same pattern of results for the Overall Performance scale *(Social Hazard Cue, $p = .004$; Visual Hazard Cue, $p = .006$; Agent Hazard Cue, $p < .001$)*. Both the Visual Hazard Cue and Agent Hazard Cue ratings were significantly lower than Control for Temporal Demand *(Visual Hazard Cue, $p = .033$, Agent Hazard Cue, $p = .01$)* and Effort *(Visual Hazard Cue, $p = .016$, Agent Hazard Cue, $p = .008$)*. The Agent Hazard Cue condition was rated significantly lower for the Mental Demand *($p = .006$)* and Frustration *($p = .002$)* sub-scales. There were also significantly lower scores in the Social Hazard Cue than the Agent Hazard Cue condition. There were no significant differences with the Physical Demand subscale and no other comparisons were significant (see Appendix F for a full list of comparisons).

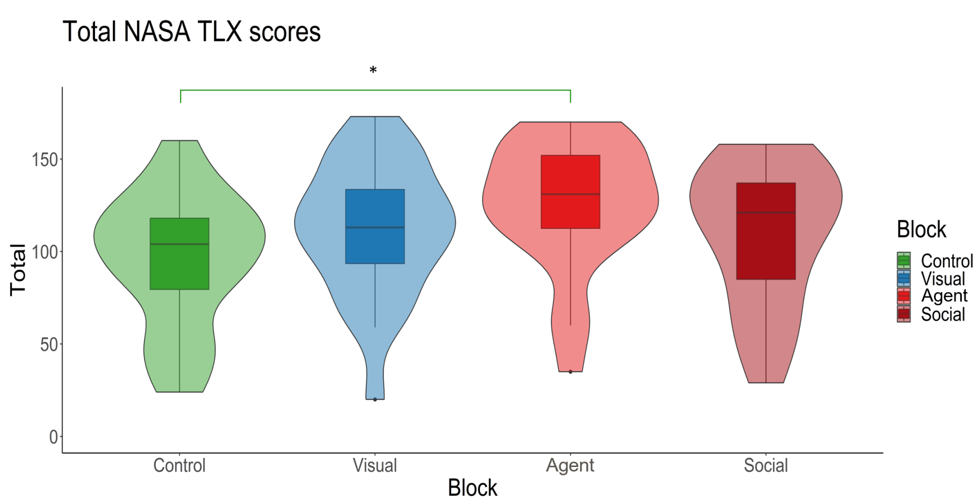


Figure 6.17: Average total NASA TLX ratings for each cue condition in Experiment 6.

NDRT Performance

A one-way repeated measures ANOVA was conducted on performance of the NDRT. However, there were no significant differences *($F(1, 70) = 0.01$, $p = .919$)* between average score in any of the conditions.

RoSAS

A series of one-way repeated measures ANOVAs were conducted comparing ratings on the RoSAS scale between the Visual and Agent Hazard Cue conditions. For the Warmth factor, there were significant differences between ratings on the Social *($F(2,46) = 11.23$, $p < .001$, $\eta^2 = 0.13$)*, Compassionate *($F(2,46) = 10.48$, $p < .001$, $\eta^2 = 0.1$)*, & Emotional *($F(2,46) = 16.21$, $p < .001$, $\eta^2 = 0.24$)* where both the Agent and Social Hazard Cue conditions were rated higher than the Visual, but with no significant differences to each other (see Table F.2 for all comparisons). For the Competence factor, there was only a significant difference on the Reliable sub-scale *($F(2,46) = 4.06$, $p = .024$, $\eta^2 = 0.04$)*, where the Agent Hazard Cue was rated lower *($p = .017$)* than the Visual Hazard Cue. There were no differences for any of the Discomfort ratings.

Rankings

Participants were again asked to rank their preferred condition for completing the driving task in order of most to least favourite. The Control condition with no non-driving task was ranked the highest, followed by the Visual Hazard Cue condition, the Agent Hazard Cue condition with the Social Hazard Cue condition being the least preferred. See Table 6.12 for the full list of rankings and Figure 6.18.

Condition	1st	2nd	3rd	4th	Ranking
Control	72	0	2	4	**78**
Visual Hazard Cue	4	39	10	4	57
Agent Hazard Cue	8	24	12	7	51
Social Hazard Cue	8	6	22	8	44

Table 6.12: Participant rankings of each Cue condition with the number of points gained for each ranking. Participants were asked to rank the conditions in order of preference for being able to complete both the driving and non-driving tasks.

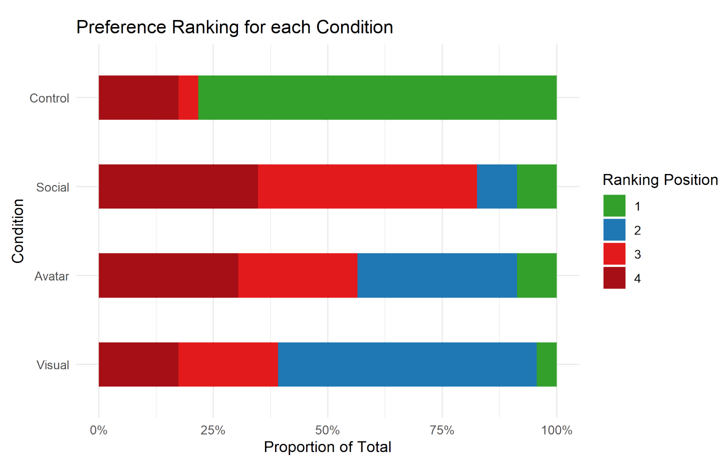


Figure 6.18: Preference ranking for each of the Conditions in Experiment 6.

Interviews

Participants were given a short post-experiment interview after completing the 4 conditions. When asked if they perceived the agent in the Social Cue condition as demonstrating an emotion, the majority of participants described variations on negative emotions such as stress, anxiety or fear. Five described the agent as having emotions, but could not define it as a single term. Nine stated they did not perceive it as having any emotional response (see Table 6.13).

Did the Social Agent express an emotion?	Count
No	9
Yes (undefined)	5
Stress	2
Anxiety	2
Fear	2
Nervous	1
Alarm	1
Danger	1
Worry	1

Table 6.13: A summary of participant responses when asked to describe what emotion the animated Social Cue condition was expressing.

Eyetracking

The same method for collecting and analysing eye tracking data as Experiment 5 was followed (see Figure 6.4). A total of 13,076,422 gaze points were recorded across all participants in all four conditions before processing. After processing, 838,670 valid fixations for analysis from 23 participants remained.

Total Fixation Count

A LME model was fitted comparing the mean number of fixations for each cue condition with participants as a random effect, with substantial explanatory power (conditional $R^2 = 0.83$). Overall, there were a significantly lower number of fixations in the Control condition compared to Visual Hazard Cue *(Est. = 108.51, t(68) = 6.393, p < .001),* Agent Hazard Cue *(Est. = 99.79, t(68) = 5.959, p < .001)* and Social Hazard Cue *(Est. = 116.71, t(68) = 6.969, p < .001)* . No other comparisons were significant (see Table 6.14 and Figure 6.19).

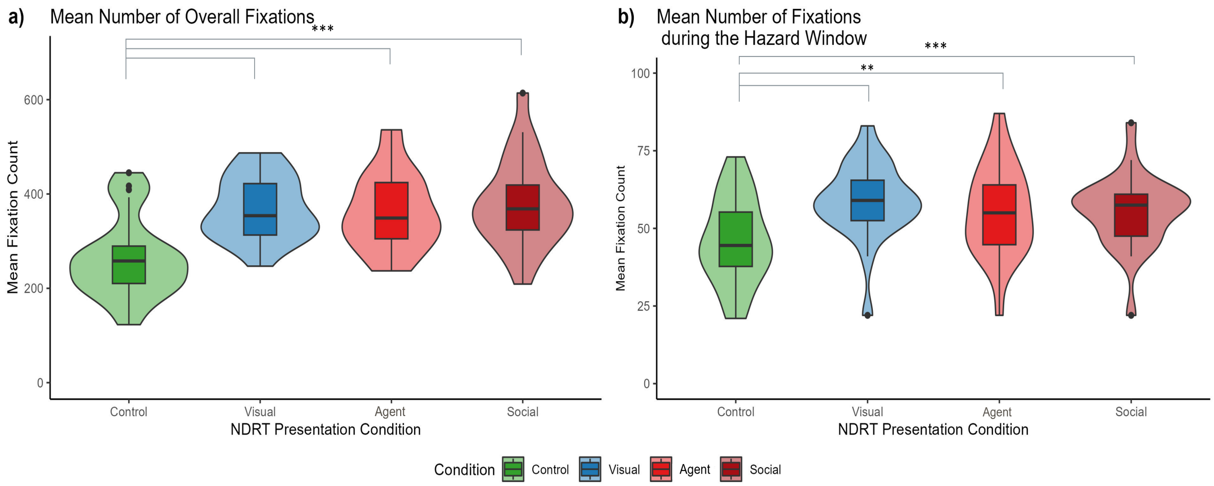


Figure 6.19: a) The mean number of fixations throughout the whole experiment for each condition and b) the mean fixation count in the last 4 seconds before the hazard appeared score for Experiment 6.

Hazard Fixation Count

A LME model was fitted comparing the mean number of fixations during the Hazard window for each cue condition as the random effect of participants was found to be minimal, with substantial explanatory power (adj. $R^2 = 0.74$). There were significantly higher number of fixations in the Visual *(Est. = 13.031, t(68.278) = 4.618, p < .001)*, Agent, *(Est. =9.458, t(68.025) = 3.397, p = .001)*, and Social *(Est. = 9.375, t(68.0.25) = 3.367, p = .001)* conditions when compared to the Control condition. No other comparison produced significantly different results (see Table 6.14 and Figure 6.19).

Correct Hazard Fixation Count

'Correct' Fixations were classified depending on whether they were located in the same area of the screen as the hazard in the video clip, during the period up to four seconds before the end of the Hazard Prediction clip. This was the length of time that the cues were evident in the AR conditions. A LME model was fitted to predict the average number of fixations in the correct AOI for each Condition. The model included participant as a random factor and had substantial explanatory power (conditional $R^2 = 0.45$), but found there were no significant differences between the mean number of correct fixations in any of the conditions.

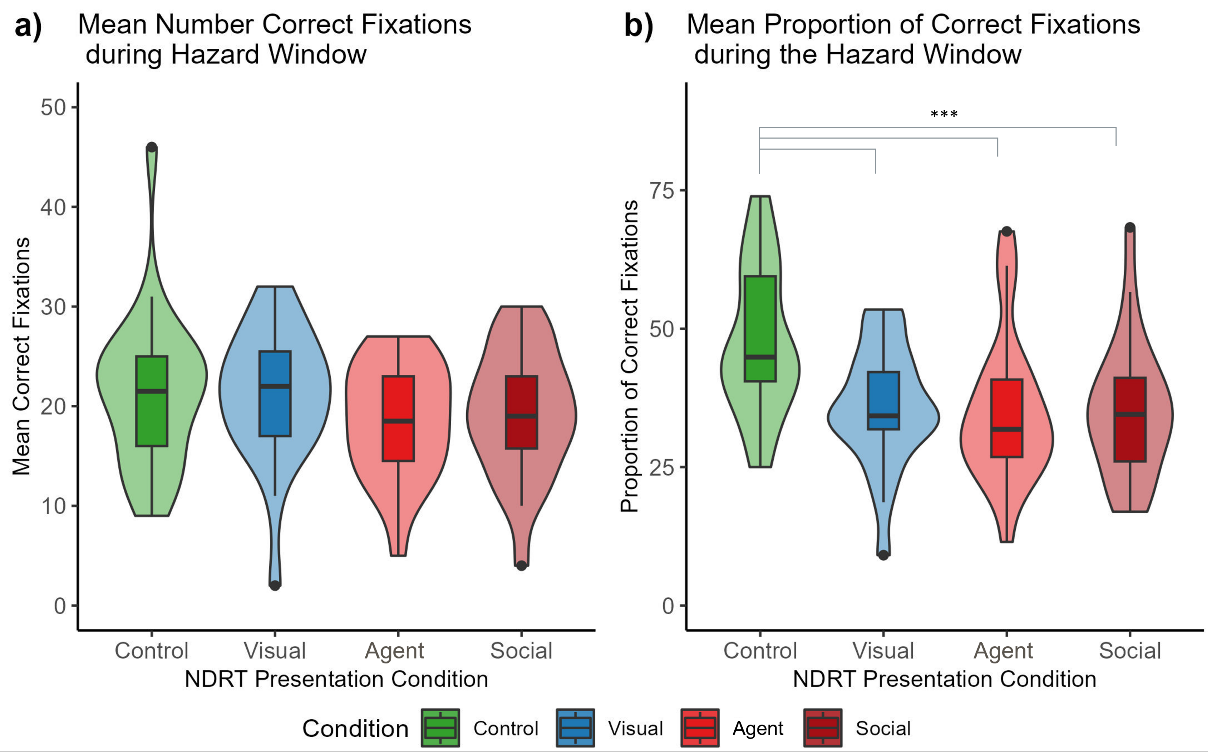


Figure 6.20: a) The mean fixation duration throughout the whole experiment for each condition and b) the mean fixation duration in the last 4 seconds before the hazard appeared score for Experiment 5.

Condition	Total Fixation			Hazard Window			Correct Fixations		
	Count	Average	SD	Count	Average	SD	Count	Average	SD
Control	6322	263.42	84.12	1103	45.96	12.77	513	21.36	7.76
Agent Hazard Cue	8717	363.21	81.05	1330	55.42	13.77	445	18.54	5.75
Social Hazard Cue	9123	380.13	93.25	1328	55.33	11.7	462	19.25	6.39
Visual Hazard Cue	8498	369.48	65.97	1350	58.7	12.33	487	21.17	6.69

Table 6.14: Summaries of the total number of fixations in each condition overall, during the Hazard Window and the number of correct fixations during the hazard window in Experiment 6.

There were significantly higher proportion of Correct fixations in the Control condition compared to the Visual *(Est. = -11.753, t(68.485) = -4.090, p < .001)*, Agent *(Est. = -12.625, t(68.171) = -4.451, p < .001)* and Social *(Est. = -12.458, t(68.171) = -4.392, p < .001)* conditions. However, there were no significantly different results from other comparisons (see Table 6.14 and Figure 6.20).

Total Fixation Duration

Due to the non-normal distribution of the data, fixation duration was z-scored and 16,576 values above 3 standard deviations from the mean duration were removed. Similarly, 4468 fixations which did not have a registered target or AOI target were removed, since the lack of a relevant gaze point target meant that the participant was looking at neither the screen nor any of the AR content.

A LME model was fitted to model the effect of Condition on average fixation duration with participant as a random effect that had substantial explanatory power (conditional $R^2 = 0.39$). In the Control condition, there were significantly longer fixation durations compared to the Visual *(Est. = -400.1, t(67.93) = -8.192, p < .001)*, Agent *(Est. = -449.74, t(67.93) = -9.209, p < .001)* and NDRT only *(Est. -425.34, t(68.11) = –8.591, p < .001)* conditions. No other comparison produced significantly different results (see Table 6.15 and Figure 6.21).

Hazard Fixation Duration

As before, 'correct' fixations were classified depending on whether they hit the same AOI area of the screen as the hazard up to four seconds before the end of the Hazard Prediction clip. A LME model was fitted to predict fixation length based on Condition with participants as a random effect, which had substantial explanatory power (conditional $R^2 = 0.29$). Within this model, there was a significant difference between the fixation duration in the Control condition and all other conditions, with fixations being shorted in the Visual *(Est. = -596.59, t(68.16) = -5.608, p < .001)*, Agent *(Est. = 409.8, t(68) = -3.905, p < .001)* and Social *(Est. = -454.56, t(68) = -4.332, p < .001)* conditions. However, there were no significant differences between any of the other conditions (see Table 6.15 and Figure 6.21).

Time to First Fixation

A LME was fitted to model the effect of Condition on time until the first fixation on the correct area of the screen. There were significantly slower first fixations in the Visual Hazard Cue *(Est. = 0.214, t(1871) = 2.222, p = .0264)* and Agent Hazard Cue *(Est. = 0.2143, t(1901) = 2.069, p = .039)*, but no difference with the Social condition. However, there were no significant differences between any of the other conditions (see Table 6.15 and Figure 6.22).

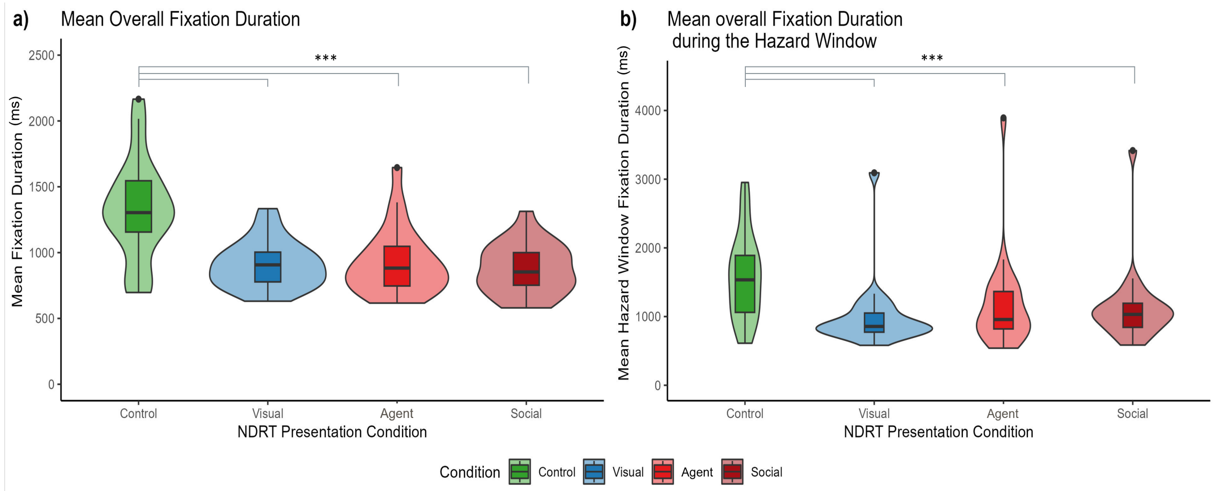


Figure 6.21: a) The mean fixation duration (ms) during the whole experiment for each condition and b) the mean fixation duration (ms) in the last 4 seconds before the hazard appeared in Experiment 6.

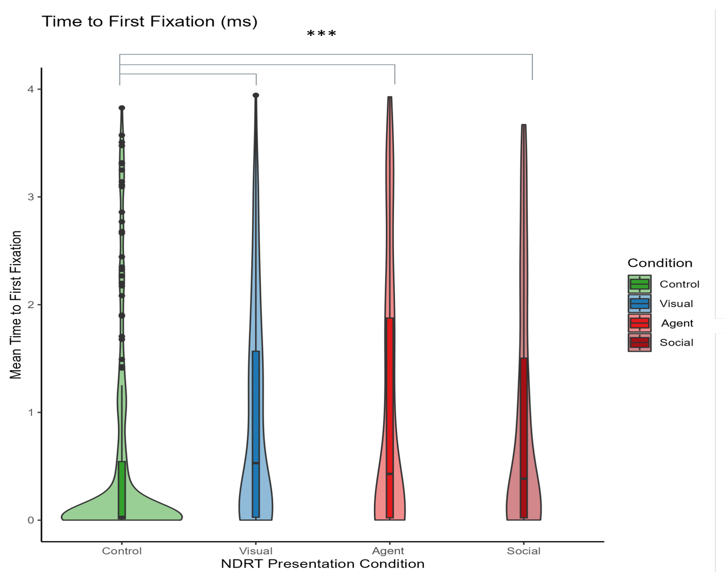


Figure 6.22: Average Time to First Fixation on the hazard for each condition in Experiment 6. The Control condition graph tending towards 0 indicates that the first fixations of participants in this group were before the 4 seconds Hazard window when the cue appeared in the other conditions, anticipating the hazard before the cues would have been shown in the AR display.

Condition	Mean Fixation Duration (ms)		Mean Hazard Fixation Duration (ms)		TTFF	
	Mean	SD	Mean	SD	Mean	SD
Control	1323.48	376.94	1415.49	1799.47	1945.2	1644.39
Agent Hazard Cue	923.38	248.06	1029.16	1263.55	2157.69	1599.99
Social Hazard Cue	873.74	190.62	1032.47	1171.18	2093.78	1589.8
Visual Hazard Cue	908.21	117.81	915.4	1003.12	2167.64	1556.2

Table 6.15: Mean fixation durations (ms) in each condition overall, mean durations during the Hazard Window and the Time to the First Fixation (TTFF) in Experiment 6.

6.8 Experiment 6 Discussion

The results from Experiment 6 provide a contrast to those of Experiment 5, as there were no differences between hazard prediction scores during an NDRT in the Visual Hazard Cue, Agent Hazard Cue conditions compared to the Control condition. It suggests that including colour change as a hazard alert to indicate danger was successful in aiding drivers to predict what happened next in the clips, even while they were performing a distracting NDRT. This was evident in both the Visual Hazard Cue and Agent Hazard Cue conditions. Furthermore, using an agent was as effective as a typical visual attentional cue, while being rated as more social. On the other hand, having the agent gesticulating towards the area of the screen in the Social Hazard Cue condition impaired hazard prediction ability compared to the other conditions. This is despite the fact that there were no differences between the confidence, subjective attention or TLX ratings between the two agents; nor were there any differences in RoSAS ratings between the two. It would indicate that the colour-change cue in the Agent Hazard Cue condition was the more useful hazard cue that alerted drivers to the presence of a hazard in the clips. Nonetheless, it is important to note that the Agent Hazard Cue condition employed the head-turning behaviour as a social cue, and drivers were still able to use the orientation of the agent's head as a social cue to infer the location of the hazard. This suggests that simpler social cues combining a hazard alert can be beneficial to distracted drivers, but more complicated social animations do not aid attention in the same way.

The eye-tracking data here follow a similar pattern to Experiment 5. There were fewer longer fixations in the Control condition compared to the NDRT conditions. While there were no differences in the number of correct fixations, the time until the first fixation was significantly faster in the Control condition as well. Like in Experiment 5, this tended towards 0 in the Control condition as this was what the bounds of the Hazard window were defined as. Despite this delay to the first fixation, the Visual and Agent Hazard Cue conditions saw no difference in hazard prediction performance, suggesting that the cues were able to convey information about the road scene in an efficient manner. This is even though the rest of the eye tracking data indicate that fixation patterns were disrupted compared to when watching the road without distraction. Conversely in the Social Cue condition, this same pattern of gaze behaviour resulted in impaired hazard prediction. These differences suggest that the combination of the positional and colour change information was sufficient to aid drivers in their hazard prediction, even though their attention towards the road and gaze behaviour was disrupted.

6.9 Chapter 6 Discussion

The two studies outlined in this chapter assess the impact of using a virtual agent to present attentional cues to aid the situational awareness of distracted drivers. The results indicate that drivers were able to use the head turning behaviour of a virtual agent to inform their hazard awareness, but only when a hazard warning cue was included in the design and not with an animated social cue. These experiments help address the research questions outlined above in the following ways:

- **RQ1)** *Is hazard prediction performance of a distracted driver affected by including visual or social cues showing positional information alone?*

  Hazard Prediction performance was significantly poorer in the NDRT conditions with positional cues compared to watching the road without distraction. This was regardless of whether the cue was a visual or a social cue, neither of which showed any significant differences compared to when the cues were absent. Overall, despite reliably indicating the location of the hazard, the positional cues provided no added benefit for hazard awareness while drivers were distracted by an NDRT presented in an AR HUD.

- **RQ2)** *Is hazard prediction performance of a distracted driver affected by including visual or social cues showing positional information with a hazard alert?*

  Including a colour-change hazard warning alert into the attentional cues maintained hazard prediction at the same level as when watching the road without any distraction. This was seen in both the visual and agent cue conditions while drivers were engaged with a distracting NDRT. These results suggest that including salient attention capturing elements in a positional cue can enhance driver awareness of the road, even if they are distracted by an NDRT which takes attention away from the driving task. This provides evidence that it is possible to use attentional cues in AR to aid drivers, even when their attention is not on the driving task. This would allow drivers to engage with an NDRT and be better prepared to resume control of the vehicle should they need to them to, without sacrificing the benefits of automated driving. Beyond simply reacting to an alert, these cues benefited the higher prediction level of situational awareness and aided drivers in predicting what happened next in the road clips, a key aspect for making safe driving decisions.

- **RQ3)** *Are distracted drivers better able to use attentional cues presented through a social agent to aid their hazard awareness compared to a visual cue?*

  Drivers were able to use the implicit head turning movements of the virtual agent to aid their hazard prediction performance, which was at similar levels to both when they were not distracted with an NDRT. This suggests that drivers can use information presented in a social modality to inform their awareness of the road. However, this was only evident when the agent was augmented with a colour change cue. Performance was significantly worse when a more animated social agent displaying the social cues was included, suggesting that an ambiguous or complicated cue creates too much of a distraction.

The results from this chapter show that using cues which provide positional information alone does not aid hazard prediction if there is nothing to signal the driver about an impending hazard, regardless of the modality of the cue. However, positional cues that changed colour to indicate danger maintained hazard prediction scores for distracted drivers the same as when there was no NDRT. While Experiment 5 showed that this NDRT significantly disrupted hazard prediction, including the colour change cues into this NDRT resulted in performance which was not significantly different to the Control condition in Experiment 6. It suggests that the combination of positional cues and hazard alerts in an AR HUD was able to provide enough information to a distracted driver to maintain their situational awareness. This is significant, as previous work has found that presenting NDRTs via an AR HUD impaired situational awareness (Radlmayr et al., 2018), even if a dynamic cue was included (Goodge et al., 2024). The findings presented here indicate that, with the appropriate cue design, it is possible to aid driver awareness during an NDRT. This is at a higher level than reacting to a TOR, but being able to accurately predict what happens next, which is key to informing their safe driving decisions (Endsley, 1995a; Crundall, 2016). It suggests that designing an AR HUD which includes these types of cues can allow a driver to engage with an NDRT in partially automated vehicle, but maintain their ability to react safely to a hazard. Furthermore, drivers were able to infer positional information from an agent with a colour change cue to aid their hazard prediction. While performance was the same with a visual cue when aided by the social cues, the agent was rated as significantly more social than a visual cue. This suggests that social cues can be an effective way of communicating information to drivers beyond using conversational interfaces (Politis et al., 2014a). In addition to the benefit of providing information from a conversational agent (Wang et al., 2022a), the results here indicate that social behavioural cues can also aid driver awareness, in particular with a distracted driver.

This suggests that virtual agents showing these social cues could be used to aid drivers in partially automated vehicles, beyond just improving interaction with the vehicle, and can have direct benefits on a driver's ability to monitor the road during an automated drive.

Limitations and Recommendations for Future Research

This chapter is an initial attempt at using a virtual agent to cue hazards in an automated driving context. Previous studies have measured the effect of characterising human-vehicle interaction in a social manner, but few have measured the direct effect of using these interactions to aid driver awareness of the road. While this chapter provides a useful contribution to the field, there are some methodological limitations.

Choice of Non-Driving Task

The gem-popping task was specifically designed to disrupt normal driver attention, as shown by the eye-tracking data from Chapter 5. Similar results were shown here, with hazard prediction scores in the NDRT only conditions of Experiment 5 significantly impacted compared to performing the task without distraction. However, it does not necessarily represent all of the NDRTs that driver might engage with during an automated drive. Consumers report wanting to engage with NDRTs including socialising, leisure activities and productivity tasks (Panagiotopoulos and Dimitrakopoulos, 2018), all of which have distinct and complex requirements of attention which the task used here does not fully represent. However, since the aim of this chapter was to investigate the impact of an NDRT on *visual attention* on visual-based Hazard Prediction task, the gem-popping NDRT was selected to represent a visual-demanding game. Further research should investigate how the social and visual cues affect hazard prediction on a wider range of NDRTs, and how interacting with different modalities of NDRT impacted the effectiveness of these cues.

Location of Cues

All of the attentional cues used in this chapter were displayed at the same controlled location for each driver to ensure they were visible in the limited field of view of the HoloLens 2. Previous work suggests that passengers have different preferences for the location of augmented reality interfaces depending on the environment (Medeiros et al., 2022), and other research looking at virtual agents typically place them in the position of a passenger (Wallbridge et al., 2022; Wang et al., 2022a). This placement was also chosen to create a realistic interaction with the agent in these studies. Doing so created a unique challenge in designing social cues which were visible from behind to facilitate this. The agent used here was intentionally simplistic, with focus on creating easily observable social cues, rather than designing a full social interaction, to compare with a visual cue.

Most prior research investigating social cues relies on face-to-face interaction using eye gaze behaviour (Admoni and Scassellati, 2017), which is a reliable method of establishing joint attention. While this was not possible in these studies, previous work suggests that head-turning can also be used to establish joint attention as an approximation for eye movements (Sheikhi and Odobez, 2015; Rubio-Fernandez et al., 2022). The agents in both studies were rated as significantly more social than the visual cues on the RoSAS scales, suggesting that the behaviour created the idea of the agent providing social cues. However, further research should evaluate different methods of presenting these types of social cues via virtual agent and how different visualisation methods affects the perceived 'social-ness' of the virtual agent. For example, placement of the virtual agent as a passenger in the periphery, in front so that cues are visible to the driver without having to take their eyes off the road, or displaying more obvious representations of gaze, such as by Trösterer et al. (2019).

Interaction with the Virtual Agent

Further to this, the level of interaction between the driver and the agent also requires consideration. The agents used here did not have any interaction with the drivers beyond displaying the social cues. However, more complex two-way interactions with a social agent have found benefits as in-vehicle displays (Karatas et al., 2016, 2018; Wang et al., 2022a). While the agents used in the studies in this chapter were perceived as more social than the visual cues based on the RoSAS ratings, there was no opportunity for interaction with them beyond observing their movement. Furthermore, there were no differences in Competence RoSAS ratings compared to the visual cues, which may have affected the propensity for participants to attend to the cues from the agents. Though previous work has suggested that abstract avatars are suitable when observing interactions (Mathis et al., 2021), trust plays a significant role in these interactions (Sun and Botev, 2021), with previous work showing that virtual agents are rated as less trustworthy than human or robot partners, (Pan and Steed, 2016), and anthropomorphism affecting perceptions of credibility (Nowak and Rauh, 2008). Future work should explore how a two-way interaction between a driver and a virtual agent, combining the social cues here with informative conversation alerts from Wang et al. (2022a) could assist driver awareness, considering how agent design affects trust in the system (Sun and Botev, 2021). Finally, the Hazard Prediction test used here as a means to empirically measure situational awareness, is not necessarily reflective of awareness throughout an entire drive, as it only measures the moment the driver disengages from the NDRT and turns their attention towards the road (Stage 3 in Janssen et al. (2019)'s model). Further research should measure driving performance in an extended or interactive setting to evaluate the effect of using social attentional cues during a realistic automated drive.

Avatars in Cars: A Benefit or a Burden for Attention?

Prior to this chapter, Wang et al. (2022a) found that using a conversational voice agent providing assistance through speech improved situational awareness scores on a SAGAT test. Combined with the findings of the studies presented here, these indicate that social information presented through an embodied virtual agents could be an effective way of informing drivers about the road and helping them maintain situational awareness in automated vehicles. This could be in addition to the visual displays that guide attention that are typically seen in AR HUDs. An embodied virtual assistant which indicates the vehicle's current state and information about the road around it, using speech and social behaviour, e.g., head turning, could be beneficial for maintaining the awareness of a driver who is distracted by an NDRT, should they need to resume control.

This would expand on the idea proposed by Janssen et al. (2019) for sharing the responsibility of attending to the road with the vehicle from a conceptual model into a social interaction. Schuß et al. (2024) showed high levels of acceptability and trust for a digital co-pilot in automated vehicles across different cultures, which indicate that facilitating a social interaction between drivers and automated vehicles is feasible. As discussed previously, characterising interaction with the vehicle in a social manner improved ratings of confidence (Wang et al., 2022b), likeability (Dong et al., 2020), and increased trust and reduced blaming the vehicle for errors (Wallbridge et al., 2022). Furthermore, speech-based agents have been shown to reduce anger levels and improve situational awareness (Jeon et al., 2015). A mixed-reality embodiment of a virtual agent would allow the benefits of these social interactions for explaining the actions of an automated vehicle. However, the findings from the studies presented here suggest that this can be taken further. Social behavioural cues can be used to impart information about the road itself to drivers when they are distracted and provide attentional alerts about the road. This could expand the role of virtual agents beyond facilitating interactions with in-car interface to more of a co-pilot position, which the driver can interact with socially and share responsibility for maintaining attention to the road with, responding to the assistant's reactions to the road.

However, caution must be taken when considering the responsibility for the vehicle when introducing virtual assistants as an in-vehicle interface. In follow-up interviews for Experiment 6, participants revealed that they outsourced the hazard prediction task to the cues, instead focusing on completing the NDRT. Only once the cue turned red did they report switching their attention to the driving task. This does not represent the shared responsibility for watching the road that Janssen et al. (2019) propose, but an offloading of responsibility onto the virtual agent. This does raise the issue of sharing responsibility for the vehicle with a social agent and how much trust a driver places in the vehicle.

Drivers exposed to automated driving systems have higher levels of trust (Körber et al., 2018), and are more prone to drowsiness (Kundinger et al., 2019) and to look away from the road (Körber et al., 2018). Regarding in-vehicle virtual agents, drivers have been shown to place more trust in an agent that looks like them (Verberne et al., 2015), and a vignette study by Hong et al. (2021) found that people would give more praise to an AI-controlled vehicle than a human driver, something which has been suggested to be moderated by personality and desirability of control (Sela and Amichai-Hamburger, 2023). Aroyo et al. (2021) describe how anthropomorphism can lead to humans overtrusting robots if they believe they are sentient. This could lead to significant issues in an automotive context, with drivers believing their automated vehicle to be more competent as it is represented by a social agent. While previous work has shown how having passengers in the vehicle leads to a reduction in unsafe driving behaviours (Michel and Meyers, 2004; Vollrath et al., 2002), it is not clear how this would transfer into an automated vehicle; would a driver feel the same level of responsibility towards a virtual agent, and how would this affect their driving performance or trust in automation?

It is important to note that this chapter does not make the assertion that social cues are superior to visual ones. Admoni and Scassellati (2012) found that performance was worse in a counterpredictive cueing task (where cues point to the opposite direction of the target) when using an avatar to cue attention compared to human facial or visual arrow cues. Though cueing was confirmatory in these studies, using an agent to indicate the location of the hazard through position alone had no benefit, suggesting that there are complex socio-cognitive factors which need to be considered when designing social cues for agents (Admoni and Scassellati, 2017). This may also be reflected in the significant increase in NASA TLX scores, which were significantly higher in the Agent Hazard Cue condition in Experiment 6, despite there being no differences in Hazard Prediction scores. It was only with the inclusion of the colour change cues in Experiment 6 that there was a benefit, a consistent finding with the literature looking at cueing attention (Pan, 2010; MacKay and Ahmetzanov, 2005; Harris et al., 2015).

Though drivers were able to use the implicit social cues, it appears that the significant factor aiding their awareness was the changing of colour of the attentional cues. The Visual Hazard Cue and Agent Hazard Cue conditions both signalled a hazard through changing the colour of the cue and showed similar levels of performance as the Control condition with no NDRT. This is likely the most important factor that aided hazard prediction here as well, signalling to participants when to switch from the NDRT to the driving task (Janssen et al., 2019). However, the key finding is that in Experiment 6, participants were able to *infer* a hazard location based on head-turning social cue paired with a colour change cue, and use that to aid their hazard prediction.

This was at the same rate as the visual cue which directly highlighted a specific location on the screen. These results suggest that, while not an effective cue by itself, implicit social information can be used to cue attention to specific locations, with the attention capturing colour-change cue aiding driver awareness while they are performing an NDRT to the same level as a visual cue.

Design Recommendations for Attentional Cues

The results from these two studies have implications for designers of in-vehicle interfaces that can be used to display NDRTs during automated driving. The design of interfaces to aid driver awareness while they are distracted is important while drivers are required to maintain attention to the road in Level 3 automated vehicles. From this, a list of recommendations was produced for designers of in-vehicle interfaces based on the results from these studies:

- **Neither Social nor Visual cues providing positional information alone aid situational awareness in a distracted driver**

  An attentional cue which highlights the location of danger but does not capture attention has no benefit on hazard awareness. Without a signal that there is approaching danger, the benefits of a positional cue are lost on a distracted driver performing an NDRT, regardless of the modality of the cue.

- **Colour change is an effective hazard alert**

  In both the Visual and Agent hazard cue conditions, a colour change before the hazard appeared was an effective signal for danger in distracted drivers. Combined with a positional cue, colour change can be an effective attentional cue towards danger, even when the driver is distracted by an NDRT.

- **Head turning behaviour can aid attention when paired with an active hazard alert**

  Using the head turning behaviour of an agent to signal location did provide some benefit to hazard awareness, even though the location of the danger was only implied. Socially resonant behaviour such as head-turning can be effective in conveying information, and can be a benefit to hazard awareness when combined with an attentional cue towards danger.

- **Overly complicated social animation cues are distracting**

  Despite using more direct social attentional cues, an animated agent which reacted emotionally and pointed towards the danger was more distracting than an agent that turned its head. The desire for a fully anthropomorphic virtual agent needs to be tempered with the distraction it may cause if used as an attentional aid.

6.10 Conclusion

The experiments in this chapter demonstrate that it is possible to maintain driver awareness during an NDRT at a similar level to without a distraction using attentional cues in an AR display. Experiment 5 investigated the effect of including positional cues on driver hazard prediction ability while distracted, either visually or as a social agent. However, performance on the hazard prediction task here with either a visual cue or a social cue was significantly worse than performing the task without distraction, and was no better than with no cue present. Experiment 6 on the other hand investigated the effect of including attentional cues with an active hazard alert on hazard prediction ability while distracted. Visual and virtual agent cues which indicated hazard location and changed colour to indicate danger, and an animated social agent cue which gesticulated towards the danger were compared to performing the hazard prediction task without distraction. There were no differences in hazard prediction between the no distraction condition and the colour change cues. This suggests it is possible to maintain driver awareness during a distracting NDRT at a similar level to viewing the road without distraction, but only with an attentional cue which signals when there is a hazard and its location. Presenting this cue as social information from a virtual agent provided similar benefits to a visual cue, but not when using an expressive social agent, which produced significantly worse scores than all other conditions. Addressing **RQ3** from this thesis (section 1.3), the results from this chapter show that it is possible to use AR to maintain driver awareness of the road while they are distracted by an NDRT through a combination of positional and attention capturing cues, allowing for drivers to engage with NDRTs without sacrificing their ability to respond to the road. Expanding on this, the results also indicate that drivers are able to use social cues from a virtual agent, allowing the possibility for in-vehicle agents to be used beyond an interaction mechanism as aids to driver awareness. Communicating information about the road to drivers in partially automated vehicles remains a challenge, especially if they are distracted by an NDRT. With more advanced mixed-reality in-car interfaces comes more interaction opportunities with the vehicle, such as presenting information in a social way via a virtual agent. The results from this chapter suggest that it is possible to use a social agent to aid driver awareness, on top of the benefits they can provide when used as in-car assistants.

Chapter 7

General Discussion & Conclusions

Figure 7.1: Images showing the Augmented Reality cues from each of the experiments, as well as the experimental setup that was used throughout the thesis.

7.1 Summary of Main Findings

This thesis investigated the effect of displaying non-driving tasks in augmented reality on the situational awareness of drivers. This was using measures of the higher level of driver cognition that is not typically captured in research investigating driver attention. First, the impact of an NDRT on reaction time was established in Chapter 3. Then, using the empirical measures set out in Chapter 4, the ability of drivers to anticipate hazards during an NDRT presented in different locations was evaluated in Chapter 5. Finally, different methods and modalities of cueing attention for distracted drives were compared in Chapter 6. The findings from the experiments presented in this thesis contribute towards understanding the cognitive implications of engaging with a distracting NDRT on the ability of drivers to perceive and process the road environment. These results help to address the Research Questions posed in Section 1.3 in the following ways:

- **RQ1)** *How does engaging with a distracting NDRT affect the ability to react to the road scene?*

Experiments 1 and 2 showed that when participants were engaged with a distracting task, their reaction time was slower on both a Go/No-Go and a Hazard Perception task. Experiments 3, 4, and 5 also showed that drivers' performance on a Hazard Prediction task was poorer while distracted by an NDRT compared to when they just focused on watching the road. This was also reflected in confidence and subjective attention ratings, which were significantly lower in all conditions compared to when attention was fully on the road. However, hazard prediction performance was greater than chance in all NDRT conditions, indicating that drivers were able to maintain some level of awareness of the driving task even while distracted. These results suggest that, although it is possible for drivers to maintain some awareness of the road, their ability to both react quickly to and properly process the road scene is inhibited when performing a NDRT.

- **RQ2)** *How does the location where a distracting NDRT is presented affect the ability to maintain awareness of the road scene?*

Experiments 1 and 2 found no difference in performance with different size presentation formats within the AR display, between a smaller centralised area and a wider presentation area. Both of these presentation formats had a detrimental effect on reaction time on Go/No-Go and Hazard Perception tasks. Experiments 3 and 4 showed that, contrary to their proposed benefits from the literature, using a heads-up display presentation for a NDRT did not improve the ability to predict hazards compared to a heads-down display. Even when an attentional cue was included in the design to signal a hazard, the distracting nature of the NDRT means that drivers were not able to maintain awareness.

This is significant for the implementation of AR HUDs as in-vehicle displays, as it suggests that simply displaying content at eye level of the road by itself is not beneficial for driver attention. The design of the tasks being performed in an AR HUD needs to be considered with regards to the impact they will have on drivers' cognition to avoid inattentional blindness.

- **RQ3)** *How does the design of attentional cues affect ability to maintain awareness of the road scene during a distracting NDRT?*

Experiment 3 and 4 indicated that utilising the dynamic nature of AR allows for more informative HUDs that can be used to influence driver attention. In Experiment 3, including an attentional cue into the AR HUD improved performance over a HDD, but hazard prediction was still poorer than without a distracting NDRT. Following this, Experiment 4 showed how AR displays can incorporate attentional cues that guide driver fixations toward hazardous events. However, this did not culminate in improved situational awareness of the driving task. Eye tracking data revealed that, while the design of the attentional cue used did lead drivers to fixate on the relevant area of the screen, this did not correspond to increased hazard prediction performance. These results highlight how the workload demands of an NDRT are an important aspect to consider when designing AR HUDs, and how this can impact the effectiveness of any attentional cues used to try and draw attention to the road.

Similar results were apparent in Experiment 5, demonstrating that without a specific hazard warning alert, there was no benefit of attentional cues that highlighted the location of a hazard. However, the results of Experiment 6 found that a positional cue combined with an attention capturing warning alert benefitted hazard prediction, comparable to when watching the road without distraction. These results show that it is possible to use AR to aid driver awareness of the road while they are distracted by an NDRT, through the combination of positional and attention capturing cues. These results suggest that using these attentional cues could allow drivers to engage with NDRTs in an AR HUD without sacrificing their ability to respond to hazardous road events. Furthermore, Experiment 6 also showed that this benefit can come from using social information to infer location of a hazard, to a similar level of performance as both a visual cue and when watching the road without distraction. Though a highly animated social cue did not provide the same benefits, these results indicate that in-vehicle virtual assistants showing social cues can be used as specific attentional aids as part of a social human-vehicle interface.

7.2 Contributions

The work presented in this thesis provides a number of contributions to the field:

- **Contribution 1:** Evidence consistent with previous work that engaging with a distracting NDRT has a detrimental impact on driver performance

As with previous research investigating the impact of NDRTs on driver attention, the findings in Chapter 3 provide supporting evidence that reaction time on both an inhibition control task and a hazard perception task was negatively affected when drivers were also engaged with an NDRT presented in AR.

- **Contribution 2:** A corpus of validated video clips which can be used to empirically measure hazard prediction ability

Measuring the cognitive processes behind driver attention is complicated, and requires complex measurement tools to do so. The Hazard Prediction test presented in Chapter 4 is an easily implemented empirical measure of hazard prediction ability, a key component of a driver's situational awareness, which can be used to assess a driver's ability to use pre-cursor cues to predict what happens next in a road scene.

- **Contribution 3:** Evidence that using a Heads-Up Display does not inherently benefit driver awareness

Evidence from the aviation industry describes the benefits of presenting information in a heads-up display, and there are suggestions that therefore they should be implemented in a driving setting. However, the experiments presented in Chapter 5 indicate that, by itself, using a heads-up display for an NDRT does not aid driver ability to maintain awareness of the road. Even if attention is cued and gaze is drawn to a particular location using a HUD, this does not necessarily equate to sufficient appraisal of the scene.

- **Contribution 4:** An initial exploration into the potential benefit of using social information to aid driver awareness.

Drivers were able to use social information provided via a virtual agent to aid their hazard prediction ability while distracted by an NDRT to the same levels as when using a visual cue and, importantly, when their attention was fully focused on the driving task. Though questions still remain regarding the design of a virtual agent as an in-car assistant, the experiments presented in Chapter 6 suggest that drivers are able to use social information to infer the location of danger to aid their situational awareness.

7.3 Limitations and Future Directions

Hazard Prediction Paradigm

The Hazard Prediction test provides a useful way of measuring the higher levels of a driver's SA of a road scene in an objective manner. If a driver is aware of the objects in the hazard prediction clip, they are able to correctly predict what happens next using anticipatory cues from the scene (Crundall, 2016; Ventsislavova and Crundall, 2018). However, this test only represents a snapshot moment of driver awareness in a complex driving scene, i.e., the last few seconds before a hazard occurs. The video clips in the test used here lasted no longer than a minute, and were presented in a test format, so participants were motivated to maintain their full attention to the road. This is not necessarily how a driver might maintain attention to the road during an extended automated drive (McKerral et al., 2023), and as a result are more likely to become fatigued or bored (Figalová et al., 2023; Thiffault and Bergeron, 2003). Previous research has shown that fatigued drivers on longer drivers react much slower to hazards (Ting et al., 2008) and have harsher braking responses (Mollicone et al., 2019). This is the likely application of automated driving in its current state however, such as in Germany where automated driving is limited to 60 km/h in a motorway setting (Federal Ministry for Digital and Transport, 2021). Future research should seek to employ this hazard prediction paradigm with longer hazard clip durations to measure the predictive ability of drivers after an extended time monitoring the road, or performing an NDRT.

Laboratory-Based Experiments

Further to this, measuring how drivers react in a real situation is a consistent problem when researching driver behaviour. It is not possible to capture how a driver is processing a road scenario in situ safely, nor to probe them on their ability to anticipate what happens next. The *Hazard Perception* test is widely used as a measure of a driver's ability to react to hazards on the road, while the *Hazard Prediction* test is used to probe the higher level of situational awareness that experienced drivers use to predict what happens next in the road scene. The clips used only contain one correct answer in order to provide an objective score. In reality, however, driving scenes contain a myriad of different hazards, pre-hazards and potential hazards (Crundall, 2016), as well as other road users whose presence have no impact on a driver. Nonetheless, both of these tests rely on using snapshot judgements of an isolated driving scene where there is a hazard present and a correct answer. Furthermore, there is evidence that familiarisation with an AV takes approximately 10 minutes (Omozik et al., 2019), which was also approximately the length of the conditions in the experiments presented here.

Probing driver awareness in a laboratory setting has value in that it allows the measurement of a driver's representation of the road at a particular moment in time. This is not possible during a real life drive (it is not possible to pause reality and ask a driver to predict what happens next). However, it is also important to measure the actions taken by drivers in a realistic driving setting to measure the way in which actual driver behaviour is affected. Future research should attempt to use these dual-task paradigms in realistic driving scenarios, e.g., in a driving simulator or an on-road experiment to investigate the specific effects on driver *behaviour* beyond their cognitive state.

Choice of Non-Driving Tasks

The non-driving tasks presented in this thesis were based on theoretical models of driver attention and driver gaze behaviour (or based on examples of non-driving tasks from previous literature in the case of Experiment 4). They were designed to disrupt these patterns of attention and to allow a measurement of the ability of drivers to maintain awareness while distracted, which was evidenced by the impact they had on the driving performance measures used. They provide an insight into the impact of a distracting task might have an a drivers ability to process the road. On the other hand, the list of realistic non-driving related tasks is as long as whatever a person would do to relax, and could consist of watching video content, doing work, playing games, interacting with personal devices, talking to passengers, eating, sleeping etc. Each of these NDRTs will have specific demands on different aspects of driving behaviour, from physically taking eyes off the road (Pakdamanian et al., 2022) to reducing the ability to the process the road scene even when eyes are on the road (Radlmayr et al., 2018) or in the case of sleeping, not even being conscious to react to any type of demand from the vehicle.

The visually demanding NDRTs used for the experiments presented here do not represent the full spectrum of potential NDRTs that a driver of an automated vehicle might engage with, nor the different cognitive and physical requirements these might present to a driver. Even in the visual domain, for example, watching a feature length film with visual and auditory information is likely to be much more distracting than the NDRTs used here, even presented when as a HUD (Horrey and Wickens, 2003). Additionally, though talking to passengers can provide some benefit to drivers, (Orsi et al., 2013; Drews et al., 2008), the impact of doing so on regaining control is not clear. As such, further work should seek to establish the specific cognitive requirements as well as the impact of different types and modalities of NDRTs before designing an AR interface to assist driver awareness.

Modality of AR displays

Following on from this, the NDRTs and attentional cues used in this thesis were all in the visual domain, being presented via the HoloLens 2 headset without any auditory, tactile or other modality of alert. This approach was taken for this thesis as the driving performance measure used was purely visual, with the audio removed. Plenty of research has indicated however that multimodal alerts are better for capturing driver attention (Politis et al., 2014b, 2015, 2017), while the attentional cues used here were only visual in nature. It is possible that using attentional cues in a different modality, e.g., an auditory cue signalling location, may be more beneficial than the attentional cues presented here. This was avoided here to 1) investigate the specific effect of two conflicting tasks in the visual domain, since driving is a predominantly visual task (Crundall and Underwood, 2011) and 2) avoid inherent benefits evidenced from presenting an alert in a different modality to the driving task, e.g., an auditory alert, which has been shown to have benefits on attracting driver attention (Politis et al., 2015; Geitner et al., 2019; Ward and Helton, 2022). Further work should investigate how specific interactions of modalities and different types of attentional cues can be used to alert drivers and impact situational awareness, as well as the effect of using a driving performance measure which incorporates multiple modalities.

Demographics

The participants recruited for the majority of the experiments in this thesis were recruited for their driver experience. However, given that recruitment took place through institutional channels (participants pools, university notice boards, university online forums, etc.), the samples for the experiments are limited to those who have access to these channels as well as the means, i.e., university staff, students, and affiliates. One benefit to using this sample however is that a large number of these participants were not just from the UK, and represented a broader range of driving behaviours from many different driving cultures. On the other hand, one particular limitation of the experiments here is the relatively narrow age sample, with mean ages for each experiment typically ranging between 25 and 37 years old. There is evidence that driver attention and hazard perception ability deteriorates with age (Borowsky et al., 2010; Memmert, 2006). Older drivers engage with NDRTs in automated vehicles in different fashions (Srour Zreik et al., 2023) and have specific design requirements for attentional alerts (Srour Zreik et al., 2024). Further work needs to explore how drivers of different age groups are affected when engaging with a distracting NDRT, both in general and when using an AR HUD.

7.4 Overall Discussion

In-vehicle Augmented Reality

The aim of this thesis as stated in section 1.2 was to investigate whether it is possible to use augmented reality to aid the situational awareness of drivers. The results from the experiments described in this thesis provide evidence that this is possible, but with important considerations for the implementation of in-vehicle AR HUDs and the effect it has on driver cognition. It was has previously been established that presenting driving-related information in a HUD display has benefits over a HDD (Smith et al., 2015, 2016), and so presenting NDRTs in this way is a sensible suggestion to keeping drivers in the loop (Schroeter et al., 2014; Schömig et al., 2018). However, the results presented here indicate that, for a distracted driver, a HUD does not provide the same benefit by itself when displaying an NDRT. Eye tracking data indicate that, despite fixating on the road in the hazard prediction clips, drivers were not able to sufficiently process the road scene and their hazard awareness was impaired, even when they were led to fixate on the specific area where a hazard is present. These results are consistent with the inattentional blindness literature (Simons, 2007), which demonstrates that drivers not anticipating a certain road event fail to properly process it (Crundall et al., 2012, 2017).

Few other studies investigating the impact of AR displays on driver attention measure situational awareness empirically. Many focus on driver gaze behaviour and measure, for instance, eyes-on-road time (Vicente et al., 2015; Walch et al., 2019; Karatas et al., 2020). The assumption that a driver is paying attention to the road because they are looking at it is the basis of many driver monitoring systems which utilise eye tracking (Dong et al., 2010; Koesdwiady et al., 2016), and is an important part of the understanding the perceptions of drivers (Underwood et al., 2003; Underwood, 2007; Yang et al., 2021a). However the results from this thesis suggest, in particular Experiment 4, that this does not necessarily correspond to a state of situational awareness or having fully appraised the danger. Guiding fixations to the correct part of the screen where the upcoming hazard occurs did not aid hazard prediction performance. This is a significant problem for designers of in-car AR HUDs, who must not overlook the 'Look-but-Fail-to-See' effect (Brown, 2002; Langham et al., 2002) when creating AR interfaces for displaying content to drivers. This may be a facet of the increased visual workload that is required by drivers in the experiments presented here. They were asked to perform two visually complex tasks (reacting to the NDRTs and maintaining attention to the hazard clips), the overlap of which has been shown to impair performance more generally (Brünken et al., 2002; Jackson et al., 2023) and in a driving context (Ward and Helton, 2022; Walker et al., 2021).

However, this is what is being asked of drivers of automated vehicles who need to be ready to resume control regardless of any NDRT they are engaged with. Situational awareness is an important factor to consider when evaluating how an AR interface may impact driver attention. Whether a driver is able to actually establish awareness of the road scene as Endsley (1995b) suggests is key to facilitating a smooth takeover of control, rather than simply just fixating on the hazardous area of the road but being thrust into controlling the vehicle without being suitably prepared to react to the road scene.

Attentional Cues

To attempt to mitigate this, multiple attentional cues of different designs to aid situational awareness were investigated in this thesis, making use of the dynamic nature that an AR HUD can provide to highlight important areas or road users to the driver. In Experiment 3, this attentional cue elevated performance over a HDD, but performance was still worse than when watching the road without distraction. This effect was lost in Experiment 4 however with a more distracting and demanding NDRT. Experiments 5 and 6 evaluated an attentional cue that was always visible to the driver during the NDRT. This constant positional cue on screen helped maintain Hazard Prediction ability, but only when paired with an explicit alert that indicated the presence of a hazard. The results in Experiment 5 could be an example of attenuation, where the impact of an unattended stimulus reduces over time (Yiğit-Elliott et al., 2011; Cardoso-Leite et al., 2010). However, the presence of visual attenuation is not a consistent finding, with some studies indicating that it may be a learned effect rather than an automatic one (Schwarz et al., 2018). One potential explanation for this is, as participants in Experiment 5 became used to the positional cues, their attention to them could have waned and thus they provided no beneficial information for the hazard prediction task. Without an attention capturing aspect, there was nothing to differentiate between normal operation of the cue and when it was highlighting a hazard. Additionally, participants were motivated to perform both the NDRT and the hazard prediction task, and so paying attention to the location of an apparently erratic positional cue did not provide any benefit.

When asked about their strategies for performing both tasks in follow-up interviews in these experiments, participants reported 'outsourcing' the hazard awareness task to the AR interface, only switching to that task when the cue was present. This lead to impaired performance in Experiments 3 and 4, where the attentional cues used did not supply as much information to the driver to maintain their hazard prediction performance. Similarly in Experiment 5, the positional cue alone did not provide enough information to inform drivers when to switch tasks from the NDRT to the hazard prediction task.

In Experiment 6 however, this was possible using both positional information and the hazard warning alert from the cue, in both a visual and social domain. This range of results from a variety of cues indicate that the design of these attentional cues and the methods in which they provided information to the distracted drivers were fundamental to how much aid they provided.

Sharing Responsibility for the Road

The goal of evaluating these types of attentional cues is to design interfaces which aid in the sharing of responsibility for the driving task between the driver and the AV. This would allow drivers to engage with the non-driving tasks that are proposed as the benefit of automation (Panagiotopoulos and Dimitrakopoulos, 2018), but without comprising on safety in partially automated vehicles. Typically, responsibility for the driving task has been laid out in a dichotomous format; either the computer or the human is in charge, with control handed over between the two. This is the way in which recent legislation interprets responsibility in AVs. (UK Parliament, 2024), as either User-In-Charge or No-User-In-Charge (UK and Scottish Law Commissions, 2020). This is not conducive to allowing drivers to perform an NDRT, given the evidence that attention is disrupted when doing so. Expecting a driver to remain fully attentive during the automated drive negates the possibility for any non-driving activities.

However, this binary human-computer control model does not represent a realistic way that drivers will engage with AVs. Alternatively, Janssen et al. (2019) describe a 10-step model of attention when performing an NDRT, based on that of Boehm-Davis and Remington (2009) describing the cognitive effects of interruption. Janssen et al. (2019)'s model of shared responsibility theoretically allows for greater levels of awareness to the road, in contrast to a driver-in-charge/vehicle-in-charge dichotomy. Rather than switching between tasks, this model describes multiple interleaved stages between attending to an NDRT and attending to the road. Instead of requiring a quick takeover which has been shown to be impacted by NDRTs (Dogan et al., 2019; Ou et al., 2021a), drivers are instead slowly reintroduced to the driving task in a series of steps.

The work presented in this thesis measures the situational awareness of drivers at the third stage of Janssen et al. (2019)'s model, at the point the driver disengages from their non-driving task and turns their attention towards the road (see Figure 7.2). This is the point where AR interfaces can provide the most benefit, facilitating this task-switching stage by providing additional information to drivers. What this thesis shows is that presenting information at eye-level with the road as a HUD does not provide a benefit to situational awareness at this stage, and that the demands of the NDRT impact the driver's ability to process the road.

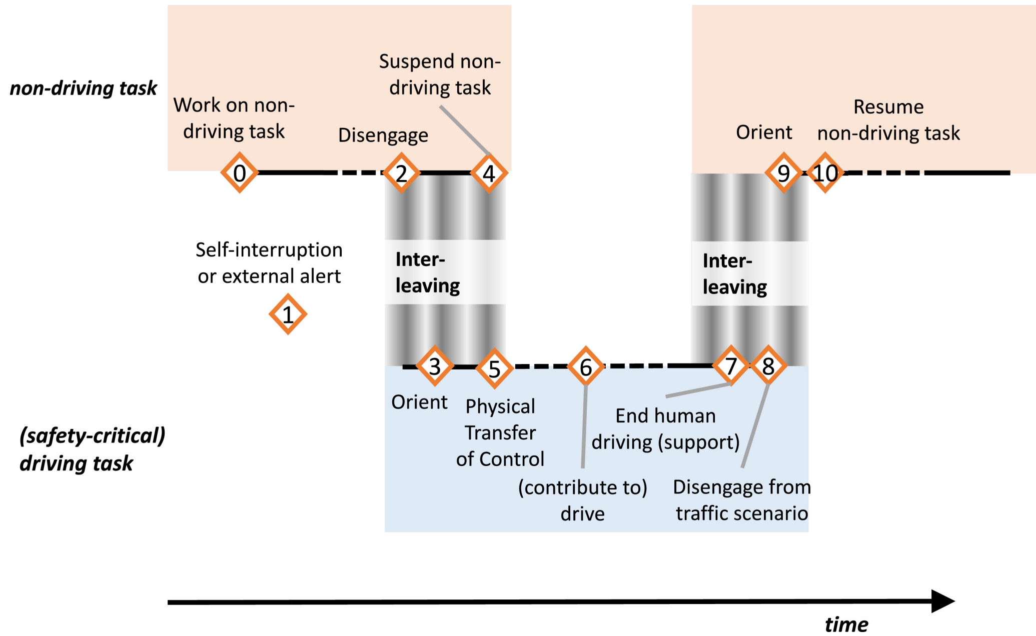


Figure 7.2: The transition of control stages as posed by Janssen et al. (2019), describing the multiple stages for a driver sharing control of an automate vehicle.

However, the results of Experiment 6 demonstrate that it is possible to design an AR interface that can provide this information to drivers even when they are distracted. Rather than as a direct application in a takeover request display, the use of the attentional cues in AR are potentially better suited for the interleaving stages of Janssen et al. (2019)'s model to facilitate driver awareness during an NDRT, should a takeover be needed. The results from this thesis demonstrate that it is possible to aid the situational awareness of drivers using an AR display, even when they are distracted by an NDRT, depending on the design of the cue and the workload of the NDRT.

In-Vehicle Assistants

Though more sophisticated AVs are not currently available to the public, there is still a significant research focus on the impact they may have on society. It may then seem even more ambitious to investigate the effects of using an in-vehicle assistant for communicating information to drivers in AVs. However, interacting with the car in a social capacity already exists through conversational interfaces, and concepts such as the in-car Omniverse Avatar concepts by NVIDIA (2021) (see Figure 7.3) indicate that incorporating virtual assistants into future vehicle interfaces has momentum. Typically these have been reserved for interacting with an interface using speech and keeping attention on the road. However, there is an increase in research investigating how to incorporate social artificial agents into these types of roles. In a driving context, there is evidence that people respond positively to virtual assistants in terms of trust (Wang et al., 2022a) and attention to the road (Karatas et al., 2016, 2018).

Figure 7.3: Concept of the NVIDIA Omniverse Concierge, an in-car virtual avatar that the driver interacts with to control the automated vehicle.

The results from Experiment 6 contribute to this in suggesting that it is possible to use social information to impart this hazard awareness as well, through head-turning behaviour of an agent. Beyond a simple voice interface, mixed reality displays could allow for more complex interactions with an in-car assistant, including using them to deliver important information about the road to a distracted driver. This could have benefits not only for driver attention, but for increasing trust in automation (Pan and Steed, 2016; Rheu et al., 2021) and enhancing the explainability of the automated driving system (Rheu et al., 2021). Rather than design a complex HMI system to do this, it leverages a modality that humans are adept at processing efficiently; information from a social interaction. While previous research has investigated using embodied robots as in-vehicle assistants (Dong et al., 2010; Wang et al., 2022a), the ways in which humans interact with robots are not simple, particularly in a driving context (Dong et al., 2020; Aroyo et al., 2021). Mixed reality interfaces, however, allow for a simpler and more dynamic implementations, such as customising the agent to enhance user perceptions and trust (Verberne et al., 2015). Though the agent used here was a simplistic one with a limited range of animation, drivers were still able to use the social cues it presented to aid their awareness. In fact, more simplistic social cues may provide more of a benefit when presented in this format (Häuslschmid et al., 2017; Wang et al., 2022a; Dong et al., 2020). Future endeavours into creating explainable autonomous systems should consider using social information to enhance the interaction experience with a system, and the potential benefits that using social information could provide by applying human-robot interaction concepts to driver attention.

Cybersickness

Any tasks presented in AR, including those presented in this thesis, could cause cybersickness symptoms to emerge if deployed in a real vehicle. Increased cybersickness has been linked to reducing attentional engagement in a VR setting (Li et al., 2021) which may impact attention to the driving task. However, studies investigating motion sickness for passengers have demonstrated how simulating motion can induce sickness symptoms (Pöhlmann et al., 2022a,b) and manipulating it can reduce their occurrence (Pöhlmann et al., 2024). Despite a focus on the potential impact of in-car VR (McGill et al., 2017, 2020; Pöhlmann et al., 2022a), there has been little research investigating how using AR could also influence cybersickness symptoms. It is not clear to what extent an AR HUD impacts motion sickness during a drive.

While this did not form part of the research focus of this thesis, the results here mimic those of Hughes et al. (2020), who found that 20 minutes of exposure to AR lead to increased sickness symptoms. Extended exposure to the AR interface here led to the emergence of sickness symptoms in the later timepoints of the experiments. There tended to be no significant differences in the change in scores over time, but distinctive differences were apparent in Experiments 2, 3, and 4 with significantly higher scores towards the end of the experiment. This was evident as increases in the Disorientation and Oculomotor discomfort subscales, but not the Nausea subscale, which Stanney et al. (1997) describes as distinctive of cybersickness. This is potentially due to the lack of physical motion that typically occurs during experiments measuring cybersickness (Duh et al., 2004; Pettijohn et al., 2020), compared to the active motion when investigating motion sickness (Pöhlmann et al., 2022a, 2024). Should in-car AR displays become a common interface for AVs however, the issue of motion sickness is likely to affect drivers in a similar way (Hughes et al., 2020), especially if they are to be used during extended automated drives. Further research looking at implementing mixed reality driver assessment tools should investigate the role of continuous vection on cybersickness levels in AR in a similar vein to that investigating in-car VR displays, and whether manipulating AR motion cues can reduce the potential onset of sickness symptoms.

7.5 Closing Remarks

This thesis evaluated the use of augmented reality to present non-driving tasks to drivers of automated vehicles. Through 6 experiments, empirical measures of driver attention and situational awareness were used to assess the impact of engaging with a distracting task while attempting to maintain attention on the road. This thesis contributes to the field by providing evidence of the impact of presenting an NDRT via AR on driving performance related measures, and indicating that a heads-up display presentation for an NDRT does not convey the same benefits found in previous research for driving-related information. However, this thesis demonstrates that with the inclusion of a positional cue with an attention-capturing hazard warning, it is possible to maintain the situational awareness of drivers at the same level as solely focusing on the road while they are engaged with a distracting non-driving task. It also provides an investigation into the efficacy of using social information to aid driver's hazard prediction, which indicate the potential of this modality for displaying information about the road to aid awareness. Overall, the findings in this thesis have significant implications for the applied transport psychology and automotive user interfaces domains. Evaluating the use of augmented reality for presenting information to drivers distracted by a non-driving task has demonstrated that it is possible to design attentional cues that can be used within these displays to aid situational awareness during an automated drive.

Bibliography

H. Admoni and B. Scassellati. Robot gaze is different from human gaze: Evidence that robot gaze does not cue reflexive attention. In *Proceedings of the "Gaze in Human-Robot Interaction" Workshop at HRI*, 2012.

H. Admoni and B. Scassellati. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction*, 6(1):25–63, 2017.

A. Al-Taie, G. Wilson, F. Pollick, and S. A. Brewster. Pimp my ride: Designing versatile ehmis for cyclists. In *15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '23, page 213–223, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701054. doi: 10.1145/3580585.3607161. URL `https://doi.org/10.1145/3580585.3607161`.

C. Y. Arias-Portela, J. Mora-Vargas, and M. Caro. Situational awareness assessment of drivers boosted by eye-tracking metrics: A literature review. *Applied Sciences*, 14(4):1611, 2024.

A. M. Aroyo, J. De Bruyne, O. Dheu, E. Fosch-Villaronga, A. Gudkov, H. Hoch, S. Jones, C. Lutz, H. Sætra, M. Solberg, et al. Overtrusting robots: Setting a research agenda to mitigate overtrust in automation. *Paladyn, Journal of Behavioral Robotics*, 12(1):423–436, 2021.

M. Asbridge, J. R. Brubacher, and H. Chan. Cell phone use and traffic crash risk: a culpability analysis. *International journal of epidemiology*, 42(1):259–267, 2013.

Axios. Cities testing self-driving driverless taxis: Waymo expands robotaxi coverage. `https://www.axios.com/2023/08/29/cities-testing-self-driving-driverless-taxis-robotaxi-waymo`, 2023. Accessed: 2024-08-07.

S. Aziz and O. Komogortsev. An assessment of the eye tracking signal quality captured in the hololens 2. In *2022 Symposium on eye tracking research and applications*, pages 1–6, 2022.

R. T. Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4): 355–385, 1997.

L. Bainbridge. Ironies of automation. In *Analysis, design and evaluation of man–machine systems*, pages 129–135. Elsevier, 1983.

K. Bark, C. Tran, K. Fujimura, and V. Ng-Thow-Hing. Personal navi: Benefits of an augmented reality navigational aid using a see-thru 3d volumetric hud. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 1–8, 2014.

D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278, 2013.

Basemark. Use cases of augmented reality in vehicles, 2024. URL `https://www.basemark.com/news/use-cases-of-augmented-reality-in-vehicles/`. Accessed: 2024-08-18.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

K. Bauerfeind, J. Drüke, J. Schneider, A. Haar, L. Bendewald, and M. Baumann. Navigating with augmented reality–how does it affect drivers' mental load? *Applied ergonomics*, 94:103398, 2021.

BBC. Highway code: Watching tv in self-driving cars to be allowed, April 2022. URL `https://www.bbc.co.uk/news/technology-61155735`.

BBC News. Highway code: Watching tv in self-driving cars to be allowed, 2022. URL `https://www.bbc.com/news/technology-61155735`.

V. Beanland, M. G. Lenné, and G. Underwood. Safety in numbers: Target prevalence affects the detection of vehicles during simulated driving. *Attention, Perception, & Psychophysics*, 76:805–813, 2014.

D. Beattie, L. Baillie, M. Halvey, and R. McCall. What's around the corner? enhancing driver awareness in autonomous vehicles via in-vehicle spatial auditory displays. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 189–198, October 2014.

N. Bigoin, J. Porte, I. Kartiko, and M. Kavakli. Effects of depth cues on simulator sickness. In *1st Intenational ICST Conference on Immersive Telecommunications & Workshops*, 2010.

L. Bobor. Sci fi drones (unity asset), 2024. URL `https://assetstore.unity.com/packages/3d/characters/robots/sci-fi-drones-90326`.

D. A. Boehm-Davis and R. Remington. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention*, 41(5):1124–1129, 2009.

D. B. Boles, J. H. Bursk, J. B. Phillips, and J. R. Perdelwitz. Predicting dual-task performance with the multiple resources questionnaire (mrq). *Human factors*, 49(1):32–45, 2007.

F. Bonato, A. Bubka, S. Palmisano, D. Phillip, and G. Moreno. Vection change exacerbates simulator sickness in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(3):283–292, 2008.

A. Borowsky, D. Shinar, and T. Oron-Gilad. Age, skill, and hazard perception in driving. *Accident analysis & prevention*, 42(4):1240–1249, 2010.

A. Bourrelly, C. Jacobé de Naurois, A. Zran, F. Rampillon, J.-L. Vercher, and C. Bourdin. Long automated driving phase affects take-over performance. *IET Intelligent Transport Systems*, 13(8):1249–1255, 2019.

V. Braun and V. Clarke. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597, 2019.

C. Braunagel, D. Geisler, W. Stolzmann, W. Rosenstiel, and E. Kasneci. On the necessity of adaptive eye movement classification in conditionally automated driving scenarios. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pages 19–26, 2016.

I. D. Brown. A review of the'looked but failed to see'accident causation factor. In *Behavioural research in road safety: Eleventh seminar*, 2002.

R. Brünken, S. Steinbacher, J. L. Plass, and D. Leutner. Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental psychology*, 49(2):109, 2002.

B. R. Burdett, N. J. Starkey, and S. G. Charlton. The close to home effect in road crashes. *Safety science*, 98:1–8, 2017.

S. Cao, Y. Qin, X. Jin, L. Zhao, and M. Shen. Effect of driving experience on collision avoidance braking: an experimental investigation and computational modelling. *Behaviour & Information Technology*, 33 (9):929–940, 2014.

M. Capallera, L. Angelini, Q. Meteier, O. Abou Khaled, and E. Mugellini. Human-vehicle interaction to support driver's situation awareness in automated vehicles: A systematic review. *IEEE Transactions on intelligent vehicles*, 8(3):2551–2567, 2022.

P. Cardoso-Leite, P. Mamassian, S. Schütz-Bosbach, and F. Waszak. A new look at sensory attenuation: Action-effect anticipation affects sensitivity, not response bias. *Psychological science*, 21(12):1740–1745, 2010.

C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner. The robotic social attributes scale (rosas) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, pages 254–262, 2017.

S. Champely, C. Ekstrom, P. Dalgaard, J. Gill, S. Weibelzahl, A. Anandkumar, C. Ford, and R. V. H. D. Rosario. Power analysis functions along the lines of cohen (1988). *CRAN*, 2016. URL `https://github.com/heliosdrm/pwr`.

L. Chandrasekaran, A. Crookes, and T. C. Lansdown. Driver situation awareness–investigating the effect of passenger experience. *Transportation research part F: traffic psychology and behaviour*, 61:152–162, 2019.

A. Chapanis. Hazards associated with three signal words and four colours on warning signs. *Ergonomics*, 37(2):265–275, 1994.

A. Chaparro, L. Groff, K. Tabor, K. Sifrit, and L. J. Gugerty. Maintaining situational awareness: The role of visual attention. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 43, pages 1343–1347. SAGE Publications Sage CA: Los Angeles, CA, 1999.

P. R. Chapman and G. Underwood. Visual search of dynamic scenes: Event types and the role of experience in viewing driving situations. *Eye guidance in reading and scene perception*, pages 369–393, 1998.

S. G. Charlton and N. J. Starkey. Driving without awareness: The effects of practice and automaticity on attention and driving. *Transportation research part F: traffic psychology and behaviour*, 14(6):456–471, 2011.

S. G. Charlton and N. J. Starkey. Driving on familiar roads: Automaticity and inattention blindness. *Transportation research part F: traffic psychology and behaviour*, 19:121–133, 2013.

I. G. Chen, D. R. Durbin, M. R. Elliott, M. J. Kallan, and F. K. Winston. Trip characteristics of vehicle crashes involving child passengers. *Injury Prevention*, 11(4):219–224, 2005.

J. Crawford and A. Neal. A review of the perceptual and cognitive issues associated with the use of head-up displays in commercial aviation, 2006. ISSN 10508414.

D. Crundall. Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention*, 86:47–58, 2016.

D. Crundall and V. Kroll. Prediction and perception of hazards in professional drivers: Does hazard perception skill differ between safe and less-safe fire-appliance drivers? *Accident Analysis & Prevention*, 121:335–346, 2018.

D. Crundall and G. Underwood. Visual attention while driving: measures of eye movements used in driving research. In *Handbook of traffic psychology*, pages 137–148. Elsevier, 2011.

D. Crundall, M. Bains, P. Chapman, and G. Underwood. Regulating conversation during driving: a problem for mobile telephones? *Transportation Research Part F: Traffic Psychology and Behaviour*, 8 (3):197–211, 2005.

D. Crundall, K. Humphrey, and D. Clarke. Perception and appraisal of approaching motorcycles at junctions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(3):159–167, 2008.

D. Crundall, E. Crundall, D. Clarke, and A. Shahar. Why do car drivers fail to give way to motorcycles at t-junctions? *Accident Analysis & Prevention*, 44(1):88–96, 2012.

D. Crundall, A. Howard, and A. Young. Perceptual training to increase drivers' ability to spot motorcycles at t-junctions. *Transportation research part F: traffic psychology and behaviour*, 48:1–12, 2017.

D. Crundall, E. Van Loon, T. Baguley, and V. Kroll. A novel driving assessment combining hazard perception, hazard prediction and theory questions. *Accident Analysis & Prevention*, 149:105847, 2021.

D. E. Crundall and G. Underwood. Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, 41(4):448–458, 1998.

A. Dam and M. Jeon. A review of motion sickness in automated vehicles. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 39–48, 2021.

N. de Oliveira Faria, C. Merenda, R. Greatbatch, K. Tanous, C. Suga, K. Akash, T. Misu, and J. Gabbard. The effect of augmented reality cues on glance behavior and driver-initiated takeover on sae level 2 automated-driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 1342–1346. SAGE Publications Sage CA: Los Angeles, CA, 2021.

J. de Winter, N. Stanton, and Y. B. Eisma. Is the take-over paradigm a mere convenience? *Transportation research interdisciplinary perspectives*, 10:100370, 2021.

J. De Wit, P. Vogt, and E. Krahmer. The design and observed effects of robot-performed manual gestures: A systematic review. *ACM Transactions on Human-Robot Interaction*, 12(1):1–62, 2023.

H. A. Deery. Hazard and Risk Perception among Young Novice Drivers. *Journal of Safety Research*, 1999. ISSN 00224375. doi: 10.1016/S0022-4375(99)00018-3.

Department for Transport, UK Government. Reported road casualties in great britain: Younger driver factsheet, 2021, 2021. URL `https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-older-and-younger-driver-factsheets-2021/reported-road-casualties-in-great-britain-younger-driver-factsheet-2021`.

Department for Transport, UK Government. Reported road casualties great britain: Provisional results 2023. `https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-provisional-results-2023/reported-road-casualties-great-britain-provisional-results-2023`, 2023. Accessed: 2024-08-07.

C. Desmet and K. Diependaele. An eye-tracking study on the road examining the effects of handsfree phoning on visual attention. *Transportation research part F: traffic psychology and behaviour*, 60: 549–559, 2019.

D. Dey, F. Walker, M. Martens, and J. Terken. Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external human-machine interfaces for automated vehicles. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*, pages 369–378, 2019.

P. Di Campli San Vito, S. A. Brewster, F. Pollick, and S. White. Thermal in-car interaction for navigation. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 510–511, 2017.

P. Di Campli San Vito, S. Brewster, F. Pollick, S. White, L. Skrypchuk, and A. Mouzakitis. Investigation of thermal stimuli for lane changes. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 43–52, 2018.

P. Di Campli San Vito, G. Shakeri, S. Brewster, F. Pollick, E. Brown, L. Skrypchuk, and A. Mouzakitis. Haptic navigation cues on the steering wheel. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.

P. Di Campli San Vito, S. Brewster, F. Pollick, S. White, L. Skrypchuk, and A. Mouzakitis. Purring wheel: Thermal and vibrotactile notifications on the steering wheel. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 461–469, October 2020a.

P. Di Campli San Vito, E. Brown, S. Brewster, F. Pollick, S. Thompson, L. Skrypchuk, and A. Mouzakitis. Haptic feedback for the transfer of control in autonomous vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 34–37, 2020b.

L. L. Di Stasi, C. Diaz-Piedra, J. M. Morales, A. Kurapov, M. Tagliabue, A. Bjärtå, A. Megias, J. Bernhardsson, S. Paschenko, S. Romero, et al. A cross-cultural comparison of visual search strategies and response times in road hazard perception testing. *Accident Analysis & Prevention*, 148:105785, 2020.

L. Dixon. Autonowashing: The greenwashing of vehicle automation. *Transportation research interdisciplinary perspectives*, 5:100113, 2020.

E. Dogan, V. Honnêt, S. Masfrand, and A. Guillaume. Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation research part F: traffic psychology and behaviour*, 62:494–504, 2019.

J. Dong, E. Lawson, J. Olsen, and M. Jeon. Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 64, pages 1033–1037. SAGE Publications Sage CA: Los Angeles, CA, 2020.

Y. Dong, Z. Hu, K. Uchimura, and N. Murayama. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE transactions on intelligent transportation systems*, 12(2):596–614, 2010.

T. Drew, M. L.-H. Võ, and J. M. Wolfe. The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological science*, 24(9):1848–1853, 2013.

F. A. Drews, M. Pasupathi, and D. L. Strayer. Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, 14(4):392, 2008.

Driving and Vehicle Standards Agency. Theory test: cars, 2023. URL `https://www.gov.uk/theory-test/hazard-perception-test`.

H. B.-L. Duh, D. E. Parker, and T. A. Furness. An independent visual background reduced simulator sickness in a driving simulator. *Presence: Teleoperators & Virtual Environments*, 13(5):578–588, 2004.

S. W. D'Arcy. Ar-hud: A heads up on the road ahead, 2022. URL `https://blog.huawei.com/2022/01/04/ar-hud-road-ahead-transportation/`.

Electronic Arts Games. Bejeweled classic, July 2021.

M. R. Endsley. Measurement of situation awareness in dynamic systems. *Human factors*, 37(1):65–84, 1995a.

M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995b.

M. R. Endsley. From here to autonomy: lessons learned from human–automation research. *Human factors*, 59(1):5–27, 2017.

M. R. Endsley. Situation awareness in future autonomous vehicles: Beware of the unexpected. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VII: Ergonomics in Design, Design for All, Activity Theories for Work Analysis and Design, Affective Design 20*, pages 303–309. Springer, 2019.

M. R. Endsley and E. O. Kiris. The out-of-the-loop performance problem and level of control in automation. *Human factors*, 37(2):381–394, 1995.

M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft. A comparative analysis of sagat and sart for evaluations of situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 42, pages 82–86. SAGE Publications Sage CA: Los Angeles, CA, 1998.

A. Eriksson and N. A. Stanton. Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors*, 59(4):689–705, 2017.

D. Ernst, S. Becker, and G. Horstmann. Novelty competes with saliency for attention. *Vision research*, 168:42–52, 2020.

M. Esterman and D. Rothlein. Models of sustained attention. *Current opinion in psychology*, 29:174–180, 2019.

R. Ezzati Amini, C. Katrakazas, A. Riener, and C. Antoniou. Interaction of automated driving systems with pedestrians: Challenges, current solutions, and recommendations for ehmis. *Transport Reviews*, 41(6):788–813, 2021.

A. Faisal, M. Kamruzzaman, T. Yigitcanlar, and G. Currie. Understanding autonomous vehicles. *Journal of transport and land use*, 12(1):45–72, 2019.

E. Fathiazar, B. Wortelen, and Y. Kühl. Using eye tracking to guide driver's attention on augmented reality windshield display. In *International Conference on Human-Computer Interaction*, pages 18–31. Springer, 2023.

Federal Ministry for Digital and Transport. Act on autonomous driving, 2021. URL `https://bmdv.bund.de/SharedDocs/EN/Articles/DG/act-on-autonomous-driving.html`. Accessed: 2024-09-18.

N. Fereydooni, O. Shaer, and A. L. Kun. Switching between augmented reality and a manual-visual task: a preliminary study. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, pages 99–103, 2019.

N. Figalová, H.-J. Bieg, M. Schulz, J. Pichen, M. Baumann, L. L. Chuang, and O. Pollatos. Fatigue and mental underload further pronounced in l3 conditionally automated driving: Results from an eeg experiment on a test track. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 64–67, 2023.

M. Firth. Introduction to automotive augmented reality head-up displays using ti dlp® technology. *Technical document, May*, 2019a.

M. Firth. Ar in your car: Hows huds are changing driving, 2019b. URL `https://www.eeworldonline.com/ar-in-your-car-how-huds-are-changing-driving/`.

R. Fischer and M. Janczyk. Dual-task performance with simple tasks. In *Handbook of human multitasking*, pages 3–36. Springer, 2022.

T. Forbes Advisor. Distracted driving statistics, 2024. URL `https://www.forbes.com/advisor/legal/auto-accident/distracted-driving-statistics/#sources_section`. Accessed: 2024-08-16.

J. Friedman, R. Tibshirani, and T. Hastie. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010. doi: 10.18637/jss.v033.i01.

A. Frischen, A. P. Bayliss, and S. P. Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4):694, 2007.

C. D. Frith and U. Frith. Social cognition in humans. *Current biology*, 17(16):R724–R732, 2007.

J. L. Gabbard, G. M. Fitch, and H. Kim. Behind the glass: Driver challenges and opportunities for ar automotive applications. *Proceedings of the IEEE*, 102(2):124–136, 2014.

J. L. Gabbard, D. G. Mehra, and J. E. Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE transactions on visualization and computer graphics*, 25(6): 2228–2241, 2018.

J. L. Gabbard, M. Smith, K. Tanous, H. Kim, and B. Jonas. Ar drivesim: An immersive driving simulator for augmented reality head-up display research. *Frontiers in Robotics and AI*, 6:98, 2019.

C. Geitner, F. Biondi, L. Skrypchuk, P. Jennings, and S. Birrell. The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario. *Transportation research part F: traffic psychology and behaviour*, 65:23–33, 2019.

H. Gelderblom and L. Menge. The invisible gorilla revisited: using eye tracking to investigate inattentional blindness in interface design. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, pages 1–9, 2018.

A. Gold. Comfort - mixed reality design guidelines, 2022. URL `https://www.motortrend.com/news/2022-mercedes-eqs-ev-hyperscreen-ces/`.

C. Gold, D. Damböck, L. Lorenz, and K. Bengler. "take over!" how long does it take to get the driver back into the loop? In *Proceedings of the human factors and ergonomics society annual meeting*, volume 57, pages 1938–1942. Sage Publications Sage CA: Los Angeles, CA, 2013.

C. Gold, M. Körber, D. Lechner, and K. Bengler. Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. *Human factors*, 58(4):642–652, 2016.

T. Goodge, V. Kroll, M. Vernon, P. Ventsislavova, and D. Crundall. A comparison of cybersickness symptoms across 360-degree hazard perception and hazard prediction tests for drivers. *Applied Ergonomics*, 97:103549, 2021.

T. Goodge, F. Pollick, and S. Brewster. Can you hazard a guess? evaluating the effects of augmented reality cues on driver hazard prediction. In *ACM CHI Conference on Human Factors in Computing Systems (CHI 2024)*, O'ahu, Hawai'i, USA, May 2024. (Accepted for Publication).

A. Gordon. California dmv suspends cruise's self-driving car license after pedestrian injury, 2023. URL `https://www.vice.com/en/article/4a3ba3/california-dmv-suspends-cruises-self-driving-car-license-after-pedestrian-injury`.

T. Groechel, Z. Shi, R. Pakkar, and M. J. Matarić. Using socially expressive mixed reality arms for enhancing low-expressivity robots. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8. IEEE, 2019.

D. Gruyer, V. Magnier, K. Hamdi, L. Claussmann, O. Orfila, and A. Rakotonirainy. Perception, information processing and modeling: Critical stages for autonomous driving applications. *Annual Reviews in Control*, 44:323–341, 2017.

A. Gugliotta, P. Ventsislavova, P. Garcia-Fernandez, E. Peña-Suarez, E. Eisman, D. Crundall, and C. Castro. Are situation awareness and decision-making in driving totally conscious processes? Results of a hazard prediction task. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2017. ISSN 13698478. doi: 10.1016/j.trf.2016.11.005.

E. C. Haas and J. B. Van Erp. Multimodal warnings to enhance risk communication and safety. *Safety science*, 61:29–35, 2014.

J. Häkkinen, M. Liinasuo, J. Takatalo, and G. Nyman. Visual comfort with mobile stereoscopic gaming. In *Stereoscopic displays and virtual reality systems XIII*, volume 6055, pages 85–93. SPIE, 2006.

P. A. Hancock. Some pitfalls in the promises of automated and autonomous vehicles. *Ergonomics*, 62(4): 479–495, 2019.

M. M. Haque and S. Washington. Effects of mobile phone distraction on drivers' reaction times. *Journal of the Australasian College of Road Safety*, 24(3):20–29, 2013.

M. Harley. Testing (and trusting) mercedes-benz level 3 drive pilot in germany, 2023. URL `https://www.forbes.com/sites/michaelharley/2022/08/02/testing-and-trusting-mercedes-benz-level-3-drive-pilot-in-germany/`. 24/11/2023.

I. M. Harms, B. R. Burdett, and S. G. Charlton. The role of route familiarity in traffic participants' behaviour and transport psychology research: a systematic review. *Transportation research interdisciplinary perspectives*, 9:100331, 2021.

A. M. Harris, S. I. Becker, and R. W. Remington. Capture by colour: Evidence for dimension-specific singleton capture. *Attention, Perception, & Psychophysics*, 77:2305–2321, 2015.

S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

R. Häuslschmid, M. von Buelow, B. Pfleging, and A. Butz. Supportingtrust in autonomous driving. In *Proceedings of the 22nd international conference on intelligent user interfaces*, pages 319–329, 2017.

A. J. Hawkins. Bmw's new curved idrive display is a 'major step' toward autonomous driving, 2021. URL `https://www.theverge.com/2021/3/15/22332131/bmw-idrive-update-curved-screen-autonomous-electric-ix`.

J. Head and W. S. Helton. Sustained attention failures are primarily due to sustained cognitive load not task monotony. *Acta psychologica*, 153:87–94, 2014.

E. Herrmann, J. Call, M. V. Hernández-Lloreda, B. Hare, and M. Tomasello. Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *science*, 317(5843):1360–1366, 2007.

H. Heuer. Dual-task performance. In *Handbook of perception and action*, volume 3, pages 113–153. Elsevier, 1996.

B. L. Hills. Vision, visibility, and perception in driving. *Perception*, 9(2):183–216, 1980.

A. Hollingworth and S. Hwang. The relationship between visual working memory and attention: Retention of precise colour information in the absence of effects on perceptual selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628):20130061, 2013.

J.-W. Hong, I. Cruz, and D. Williams. Ai, you can drive my car: How we evaluate human drivers vs. self-driving cars. *Computers in Human Behavior*, 125:106944, 2021.

M. M. Hoque, T. Onuki, Y. Kobayashi, and Y. Kuno. Controlling human attention through robot's gaze behaviors. In *2011 4th International Conference on Human System Interactions, HSI 2011*, pages 195–202. IEEE, 2011.

W. J. Horrey and J. D. Wickens. Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance. In *Driving Assessment Conference*, volume 2. University of Iowa, 2003.

W. J. Horrey, A. L. Alexander, and C. D. Wickens. Does workload modulate the effects of in-vehicle display location on concurrent driving and side task performance. In *Driving Simulation Conference North America, Dearborn, Michigan*, pages 1–20, 2003.

W. J. Horrey, C. D. Wickens, and K. P. Consalus. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, 12(2):67, 2006.

M. S. Horswill. Hazard perception in driving. *Current Directions in Psychological Science*, 25(6):425–430, 2016.

M. S. Horswill and F. P. McKenna. Drivers' hazard perception ability: Situation awareness on the road. *A cognitive approach to situation awareness: Theory and application*, pages 155–175, 2004.

M. S. Horswill, K. J. Anstey, C. G. Hatherly, and J. M. Wood. The crash involvement of older drivers is associated with their hazard perception latencies. *Journal of the International Neuropsychological Society*, 16(5):939–944, 2010.

M. S. Horswill, A. Hill, and M. Wetton. Can a video-based hazard perception test used for driver licensing predict crash involvement? *Accident Analysis & Prevention*, 82:213–219, 2015.

S. Huang, X. Zhao, D. Wei, X. Song, and Y. Sun. Chatbot and fatigued driver: Exploring the use of llm-based voice assistants for driving fatigue. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2024.

C. L. Hughes, C. Fidopiastis, K. M. Stanney, P. S. Bailey, and E. Ruiz. The psychometrics of cybersickness in augmented reality. *Frontiers in Virtual Reality*, 1:602954, 2020.

A. P. Hungund and A. K. Pradhan. Impact of non-driving related tasks while operating automated driving systems (ads): A systematic review. *Accident Analysis & Prevention*, 188:107076, 2023.

R. Häuslschmid, D. Ren, F. Alt, A. Butz, and T. Höllerer. Personalizing Content Presentation on Large 3D Head-Up Displays. *Presence: Teleoperators and Virtual Environments*, 27(1):80–106, 02 2018. doi: 10.1162/pres_a_00315. URL `https://doi.org/10.1162/pres_a_00315`.

J. Ige, A. Banstola, and P. Pilkington. Mobile phone use while driving: Underestimation of a global threat. *Journal of Transport & Health*, 3(1):4–8, 2016.

A. Ingman. The head up display concept: A summary with special attention to the civil aviation industry. *A Summary with Special Attention to the Civil Aviation Industry*, 2005.

L. Itti and P. Baldi. Bayesian surprise attracts human attention. *Advances in neural information processing systems*, 18, 2005.

K. M. Jackson, T. H. Shaw, and W. S. Helton. The effects of dual-task interference on visual search and verbal memory. *Ergonomics*, 66(1):125–135, 2023.

L. Jackson, P. Chapman, and D. Crundall. What happens next? predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics*, 52(2):154–164, 2009.

S. S. Jackson. Machine learning lecture notes: Ridge regression, 2024. URL `https://bookdown.org/ssjackson300/Machine-Learning-Lecture-Notes/ridge-regression.html#introduction-to-shrinkage`. Accessed: 2024-09-13.

M. Janczyk and L. Huestegge. Effects of a no-go task 2 on task 1 performance in dual-tasking: From benefits to costs. *Attention, Perception, & Psychophysics*, 79:796–806, 2017.

C. P. Janssen, S. T. Iqbal, A. L. Kun, and S. F. Donker. Interrupted by my car? implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies*, 130:221–233, 2019.

O. Jarosch, M. Kuhnt, S. Paradies, and K. Bengler. It's out of our hands now! effects of non-driving related tasks during highly automated driving on drivers' fatigue. In *Driving Assessment Conference*, volume 9. University of Iowa, 2017.

M. Jeon, B. N. Walker, and T. M. Gable. The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. *Applied ergonomics*, 50:185–199, 2015.

T. Jiang, Y. Wang, and R. Tang. Playing games guiding attention improves situation awareness and takeover quality during automated driving. *International Journal of Human–Computer Interaction*, pages 1–14, 2023.

C. Jing, C. Shang, D. Yu, Y. Chen, and J. Zhi. The impact of different ar-hud virtual warning interfaces on the takeover performance and visual characteristics of autonomous vehicles. *Traffic injury prevention*, 23(5):277–282, 2022.

R. Jose, G. A. Lee, and M. Billinghurst. A comparative study of simulated augmented reality displays for vehicle navigation. In *Proceedings of the 28th Australian conference on computer-human interaction*, pages 40–48, 2016.

S. Jung, J. Park, J. Park, M. Choe, T. Kim, M. Choi, and S. Lee. Effect of touch button interface on in-vehicle information systems usability. *International Journal of Human–Computer Interaction*, 37 (15):1404–1422, 2021.

B. Kapitaniak, M. Walczak, M. Kosobudzki, Z. Jóźwiak, and A. Bortkiewicz. Application of eye-tracking in drivers testing: A review of research. *International journal of occupational medicine and environmental health*, 28(6), 2015.

S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn. Arett: Augmented reality eye tracking toolkit for head mounted displays. *Sensors*, 21(6):2234, 2021.

N. Karatas, S. Yoshikawa, P. R. S. De Silva, and M. Okada. Namida: Multiparty conversation based driving agents in futuristic vehicle. In *Human-Computer Interaction: Users and Contexts: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part III 17*, pages 198–207. Springer, 2015.

N. Karatas, S. Yoshikawa, P. R. De Silva, and M. Okada. Namida: How to reduce the cognitive workload of driver. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 449–450. IEEE, 2016.

N. Karatas, S. Tamura, M. Fushiki, and M. Okada. The effects of driving agent gaze following behaviors on human-autonomous car interaction. In *Social Robotics: 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings 10*, pages 541–550. Springer, 2018.

N. Karatas, T. Tanaka, K. Fujikakc, Y. Yoshihara, H. Kanamori, Y. Fuwamoto, and M. Yoshida. Evaluation of ar-hud interface during an automated intervention in manual driving. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 2158–2164. IEEE, 2020.

M. Kaufeld, M. Mundt, S. Forst, and H. Hecht. Optical see-through augmented reality can induce severe motion sickness. *Displays*, 74:102283, 2022.

N. Kaya, J. Girgis, B. Hansma, and B. Donmez. Hey, watch where you're going! an on-road study of driver scanning failures towards pedestrians and cyclists. *Accident Analysis & Prevention*, 162:106380, 2021.

S. Kee, S. B. M. Tamrin, and Y. Goh. Driving fatigue and performance among occupational drivers in simulated prolonged driving. *Global Journal of Health Science*, 2(1):167–177, 2010.

R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

L. Kettle and Y.-C. Lee. Augmented reality for vehicle-driver communication: a systematic review. *Safety*, 8(4):84, 2022.

H. Kim and J. L. Gabbard. Quantifying distraction potential of augmented reality head-up displays for vehicle drivers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 1923–1923. SAGE Publications Sage CA: Los Angeles, CA, 2018.

H. Kim and J. L. Gabbard. Assessing distraction potential of augmented reality head-up displays for vehicle drivers. *Human factors*, 64(5):852–865, 2022.

H. Kim, X. Wu, J. L. Gabbard, and N. F. Polys. Exploring head-up augmented reality interfaces for crash warning systems. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 224–227, 2013.

N. L. Kluttz, B. R. Mayes, R. W. West, and D. S. Kerby. The effect of head turn on the perception of gaze. *Vision research*, 49(15):1979–1993, 2009.

A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel. Recent trends in driver safety monitoring systems: State of the art and challenges. *IEEE transactions on vehicular technology*, 66(6):4550–4563, 2016.

M. Körber, E. Baseler, and K. Bengler. Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics*, 66:18–31, 2018.

V. S. R. Kosuru and A. K. Venkitaraman. Advancements and challenges in achieving fully autonomous self-driving vehicles. *World Journal of Advanced Research and Reviews*, 18(1):161–167, 2023.

G. Kovács, Á. Hőgye-Nagy, and G. Kurucz. Human factor aspects of situation awareness in autonomous cars–an overview of psychological approaches. *Acta Polytechnica Hungarica*, 18(7):7–24, 2021.

V. Kroll and D. Crundall. Aperture judgement in fire-appliance drivers. *Transportation research part F: traffic psychology and behaviour*, 63:55–66, 2019.

V. Kroll, A. K. Mackenzie, T. Goodge, R. Hill, R. Davies, and D. Crundall. Creating a hazard-based training and assessment tool for emergency response drivers. *Accident Analysis & Prevention*, 144: 105607, 2020.

A. Kshirsagar, M. Lim, S. Christian, and G. Hoffman. Robot gaze behaviors in human-to-robot handovers. *IEEE Robotics and Automation Letters*, 5(4):6552–6558, 2020.

A. L. Kun et al. Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human–Computer Interaction*, 11(4):201–293, 2018.

T. Kundinger, P. Wintersberger, and A. Riener. (over) trust in automated driving: The sleeping pill of tomorrow? In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*, pages 1–6, 2019.

Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka. Museum guide robot based on sociological interaction analysis. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1191–1194, 2007.

A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017. doi: 10.18637/jss.v082.i13.

M. Kyriakidis, J. C. de Winter, N. Stanton, T. Bellet, B. van Arem, K. Brookhuis, M. H. Martens, K. Bengler, J. Andersson, N. Merat, et al. A human factors perspective on automated driving. *Theoretical issues in ergonomics science*, 20(3):223–249, 2019.

S. K. Lal and A. Craig. A critical review of the psychophysiology of driver fatigue. *Biological psychology*, 55(3):173–194, 2001.

M. Langham, G. Hole, J. Edwards, and C. O'Neil. An analysis of'looked but failed to see'accidents involving parked police vehicles. *Ergonomics*, 45(3):167–185, 2002.

S. Langlois and B. Soualmi. Augmented reality versus classical hud to take over from automated driving: An aid to smooth reactions and to anticipate maneuvers. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*, pages 1571–1578. IEEE, 2016.

S. R. Langton, R. J. Watt, and V. Bruce. Do the eyes have it? cues to the direction of social attention. *Trends in cognitive sciences*, 4(2):50–59, 2000.

D. R. Large, G. Burnett, E. Crundall, G. Lawson, and L. Skrypchuk. Twist it, touch it, push it, swipe it: evaluating secondary input devices for use with an automotive touchscreen hmi. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 161–168, 2016.

A. S. Le, T. Suzuki, and H. Aoki. Evaluating driver cognitive distraction by eye tracking: From simulator to driving. *Transportation research interdisciplinary perspectives*, 4:100087, 2020.

S. Le Vine, A. Zolfaghari, and J. Polak. Autonomous cars: The tension between occupant experience and intersection capacity. *Transportation Research Part C: Emerging Technologies*, 52:1–14, 2015.

H. Lee, S. Bang, and W. Woo. Effects of background complexity and viewing distance on an ar visual search task. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 189–194. IEEE, 2020.

E. Lehtonen, O. Lappi, I. Koirikivi, and H. Summala. Effect of driving experience on anticipatory look-ahead fixations in real curve driving. *Accident Analysis & Prevention*, 70:195–208, 2014.

G. Li, M. McGill, S. Brewster, C. P. Chen, J. A. Anguera, A. Gazzaley, and F. Pollick. Multimodal biosensing for vestibular network-based cybersickness detection. *IEEE Journal of Biomedical and Health Informatics*, 26(6):2469–2480, 2021.

N. Li, T. Misu, and F. Tao. Understand driver awareness through brake behavior analysis: Reactive versus intended hard brake. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1523–1528. IEEE, 2017.

X. Li, R. Schroeter, A. Rakotonirainy, J. Kuo, and M. G. Lenné. Effects of different non-driving-related-task display modes on drivers' eye-movement patterns during take-over in an automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 70:135–148, 2020.

P. C. Lim, E. Sheppard, and D. Crundall. Cross-cultural effects on drivers' hazard perception. *Transportation research part F: traffic psychology and behaviour*, 21:194–206, 2013.

P. C. Lim, E. Sheppard, and D. Crundall. A predictive hazard perception paradigm differentiates driving experience cross-culturally. *Transportation research part F: traffic psychology and behaviour*, 26:210–217, 2014.

P. Lindemann, T.-Y. Lee, and G. Rigoll. Supporting driver situation awareness for autonomous urban driving with an augmented-reality windshield display. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 358–363. IEEE, 2018.

K. Lipovac, M. Đerić, M. Tešić, Z. Andrić, and B. Marić. Mobile phone use while driving-literary review. *Transportation research part F: traffic psychology and behaviour*, 47:132–142, 2017.

L. Lisle, C. Merenda, K. Tanous, H. Kim, J. L. Gabbard, and D. A. Bowman. Effects of volumetric augmented reality displays on human depth judgments: Implications for heads-up displays in transportation. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 11(2):1–18, 2019.

Y.-C. Liu. Effects of using head-up display in automobile context on attention demand and driving performance. *Displays*, 24(4-5):157–165, 2003.

Y. C. Liu and M. H. Wen. Comparison of head-up display (HUD) vs. head-down display (HDD): Driving performance of commercial vehicle operators in Taiwan. *International Journal of Human Computer Studies*, 2004a. ISSN 10715819. doi: 10.1016/j.ijhcs.2004.06.002.

Y.-C. Liu and M.-H. Wen. Comparison of head-up display (hud) vs. head-down display (hdd): driving performance of commercial vehicle operators in taiwan. *International Journal of Human-Computer Studies*, 61(5):679–697, 2004b.

A. Löcken, W. Heuten, and S. Boll. Autoambicar: Using ambient light to inform drivers about intentions of their automated cars. In *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*, pages 57–62, 2016.

T. L. Louw, N. Merat, and A. H. Jamson. Engaging with highly automated driving: To be or not to be in the loop? In *8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Leeds, 2015.

S.-Y. Ma, N. R. Brady, X. Han, N.-H. Yu, and T. Yeh. Exploring mixed-reality for enhancing driver warning systems: A preliminary study on attention-shifting methods and hazard perception. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 87–92, 2023.

D. G. MacKay and M. V. Ahmetzanov. Emotion, memory, and attention in the taboo stroop paradigm: An experimental analogue of flashbulb memories. *Psychological science*, 16(1):25–32, 2005.

A. K. Mackenzie and J. M. Harris. Eye movements and hazard perception in active and passive driving. *Visual cognition*, 23(6):736–757, 2015.

A. K. Mackenzie, M. L. Vernon, P. R. Cox, D. Crundall, R. C. Daly, D. Guest, A. Muhl-Richardson, and C. J. Howard. The multiple object avoidance (moa) task measures attention for action: Evidence from driving and sport. *Behavior research methods*, pages 1–22, 2021.

D. Magazzù, M. Comelli, and A. Marinoni. Are car drivers holding a motorcycle licence less responsible for motorcycle—car crash occurrence?: A non-parametric approach. *Accident Analysis & Prevention*, 38(2):365–370, 2006.

D. Makowski, D. Lüdecke, I. Patil, R. Thériault, M. S. Ben-Shachar, and B. M. Wiernik. Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption. *CRAN*, 2023. URL https://easystats.github.io/report/.

P. Mallozzi, P. Pelliccione, A. Knauss, C. Berger, and N. Mohammadiha. Autonomous vehicles: state of the art, future trends, and challenges. *Automotive systems and software engineering: State of the art and future trends*, pages 347–367, 2019.

M. Marcano, S. Díaz, J. Pérez, and E. Irigoyen. A review of shared control for automated vehicles: Theory and applications. *IEEE Transactions on Human-Machine Systems*, 50(6):475–491, 2020.

F. Mars and J. Navarro. Where we look when we drive with or without active steering wheel control. *PLoS One*, 2012.

F. Mathis, K. Vaniea, and M. Khamis. Observing virtual avatars: The impact of avatars' fidelity on identifying interactions. In *Proceedings of the 24th International Academic Mindtrek Conference*, pages 154–164, 2021.

J. Matias, C. Belletier, M. Izaute, M. Lutz, and L. Silvert. The role of perceptual and cognitive load on inattentional blindness: A systematic review and three meta-analyses. *Quarterly Journal of Experimental Psychology*, 75(10):1844–1875, 2022.

A. Matviienko, A. Löcken, A. El Ali, W. Heuten, and S. Boll. Navilight: investigating ambient light displays for turn-by-turn navigation in cars. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 283–294, 2016.

B. Maurer, S. Trösterer, M. Gärtner, M. Wuchse, A. Baumgartner, A. Meschtscherjakov, D. Wilfinger, and M. Tscheligi. Shared gaze in the car: towards a better driver-passenger collaboration. In *Adjunct proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications*, pages 1–6, 2014.

M. McGill and S. Brewster. Virtual reality passenger experiences. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, pages 434–441, 2019.

M. McGill, A. Ng, and S. Brewster. I am the passenger: how visual motion cues can influence sickness for in-car vr. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 5655–5668, 2017.

M. McGill, J. Williamson, A. Ng, F. Pollick, and S. Brewster. Challenges in passenger use of mixed reality headsets in cars and other transportation. *Virtual Reality*, 24:583–603, 2020.

M. McGill, G. Li, A. Ng, L. Bajorunaite, J. Williamson, F. Pollick, and S. Brewster. Augmented, virtual and mixed reality passenger experiences. *User Experience Design in the Era of Automated Driving*, pages 445–475, 2022.

A. McKerral, K. Pammer, and C. Gauld. Supervising the self-driving car: Situation awareness and fatigue during highly automated driving. *Accident Analysis & Prevention*, 187:107068, 2023.

D. Medeiros, M. McGill, A. Ng, R. McDermid, N. Pantidi, J. Williamson, and S. Brewster. From shielding to avoidance: Passenger augmented reality and the layout of virtual displays for productivity in shared transit. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3640–3650, 2022.

Z. Medenica, A. L. Kun, T. Paek, and O. Palinko. Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 265–274, 2011.

S. Mehrotra, M. Wang, N. Wong, J. Parker, S. C. Roberts, W. Kim, A. Romo, and W. J. Horrey. Human-machine interfaces and vehicle automation: A review of the literature. *Accident Analysis & Prevention*, 109:18–28, 2022.

N. Meiran. Modeling cognitive control in task-switching. *Psychological research*, 63(3-4):234–249, 2000.

D. Memmert. The effects of eye movements, age, and expertise on inattentional blindness. *Consciousness and cognition*, 15(3):620–627, 2006.

N. Merat and A. H. Jamson. How do drivers behave in a highly automated car? In *Driving Assessment Conference*, volume 5. University of Iowa, 2009.

N. Merat, A. H. Jamson, F. C. Lai, and O. Carsten. Highly automated driving, secondary task performance, and driver state. *Human factors*, 54(5):762–771, 2012.

N. Merat, A. H. Jamson, F. C. Lai, M. Daly, and O. M. Carsten. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 27:274–282, 2014.

K. Merfeld, M.-P. Wilhelms, and S. Henkel. Being driven autonomously–a qualitative study to elicit consumers' overarching motivational structures. *Transportation Research Part C: Emerging Technologies*, 107:229–247, 2019.

Michel and J. R. Meyers. The influence of passengers on older drivers involved in fatal crashes. *Experimental aging research*, 30(2):205–215, 2004.

Microsoft. Comfort - mixed reality design guidelines, 2023. URL `https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort`.

Microsoft. Hololens hardware, 2024. URL `https://www.microsoft.com/en-US/hololens/hardware#document-experiences`. Accessed: 2024-09-07.

B. Mok, M. Johns, S. Yang, and W. Ju. Reinventing the wheel: transforming steering wheel systems for autonomous vehicles. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 229–241, 2017.

C. Mole, J. Pekkanen, W. E. Sheppard, G. Markkula, and R. M. Wilkie. Drivers use active gaze to monitor waypoints during automated driving. *Scientific reports*, 11(1):263, 2021.

D. Mollicone, K. Kan, C. Mott, R. Bartels, S. Bruneau, M. van Wollen, A. R. Sparrow, and H. P. Van Dongen. Predicting performance and safety based on driver fatigue. *Accident Analysis & Prevention*, 126: 142–145, 2019.

C. Moore, P. J. Dunham, and P. Dunham. *Joint attention: Its origins and role in development*. Psychology Press, 2014.

W. Morales-Alvarez, O. Sipele, R. Léberon, H. H. Tadjine, and C. Olaverri-Monreal. Automated driving: A literature review of the take over request in conditional automation. *Electronics*, 9(12):2087, 2020.

C. Moran, J. M. Bennett, and P. Prabhakharan. The relationship between cognitive function and hazard perception in younger drivers. *Transportation research part F: traffic psychology and behaviour*, 74: 104–119, 2020.

R. R. Mourant and T. H. Rockwell. Strategies of visual search by novice and experienced drivers. *Human factors*, 14(4):325–335, 1972.

J. K. Muguro, P. W. Laksono, Y. Sasatake, K. Matsushita, and M. Sasaki. User monitoring in autonomous driving system using gamified task: A case for vr/ar in-car gaming. *Multimodal Technologies and Interaction*, 5(8):40, 2021.

A. L. Müller, N. Fernandes-Estrela, R. Hetfleisch, L. Zecha, and B. Abendroth. Effects of non-driving related tasks on mental workload and take-over times during conditional automated driving. *European transport research review*, 13(1):1–15, 2021.

B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *2006 6th IEEE-RAS International Conference on Humanoid Robots*, pages 518–523. IEEE, 2006.

A. Nachiappan, N. de Oliveira Faria, and J. L. Gabbard. Can augmented-reality head-up display improve driving performance on monotonous drives? In *Proceedings of the 2021 Institute of Industrial and Systems Engineers (IISE) Annual Conference*, 2021.

K. Nachmann, B. Pillot, P. Moore, and E. Wiese. Driving with robots: Mind perception and propensity for aggressive driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 64, pages 1965–1970. SAGE Publications Sage CA: Los Angeles, CA, 2020.

L. Niu, S. Gao, J. Shi, C. Wu, Y. Wang, S. Ma, D. Wang, Z. Yang, and H. Li. Are warnings suitable for presentation in head-up display? a meta-analysis for the effect of head-up display warning on driving performance. *Transportation research record*, 2024.

B. A. Nosek and M. R. Banaji. The go/no-go association task. *Social cognition*, 19(6):625–666, 2001.

K. L. Nowak and C. Rauh. Choose your "buddy icon" carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions. *Computers in Human Behavior*, 24(4):1473–1493, 2008.

NVIDIA. NVIDIA DRIVE Concierge With Omniverse Avatar, 2021. URL `https://www.youtube.com/watch?v=7zKAKgc7Ecs`. Accessed: 2024-07-10.

B. Oktay and B. Cangöz. I thought i saw "zorro": An inattentional blindness study. *Archives of Neuropsychiatry*, 55(1):59, 2018.

L. Oliveira, C. Burns, J. Luton, S. Iyer, and S. Birrell. The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 72:280–296, 2020.

K. Omozik, Y. Yang, I. Kuntermann, S. Hergeth, and K. Bengler. How long does it take to relax? observation of driver behavior during real-world conditionally automated driving. In *Driving Assessment Conference*, volume 10. University of Iowa, 2019.

C. Orsi, P. Marchetti, C. Montomoli, and A. Morandi. Car crashes: The effect of passenger presence and other factors on driver outcome. *Safety science*, 57:35–43, 2013.

Y.-K. Ou, W.-X. Huang, and C.-W. Fang. Effects of different takeover request interfaces on takeover behavior and performance during conditionally automated driving. *Accident Analysis & Prevention*, 162:106425, 2021a.

Y.-K. Ou, W.-X. Huang, and C.-W. Fang. Effects of different takeover request interfaces on takeover behavior and performance during conditionally automated driving. *Accident Analysis & Prevention*, 162:106425, 2021b.

S. Oviatt. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 871–880, 2006.

O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Washington. Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation research part C: emerging technologies*, 72:360–380, 2016.

E. Pakdamanian, E. Hu, S. Sheng, S. Kraus, S. Heo, and L. Feng. Enjoy the ride consciously with cawa: Context-aware advisory warnings for automated driving. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 75–85, 2022.

Y. Pan. Attentional capture by working memory contents. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 64(2):124, 2010.

Y. Pan and A. Steed. A comparison of avatar-, video-, and robot-mediated interaction on users' trust in expertise. *Frontiers in Robotics and AI*, 3:12, 2016.

I. Panagiotopoulos and G. Dimitrakopoulos. An empirical investigation on consumers' intentions towards autonomous driving. *Transportation research part C: emerging technologies*, 95:773–784, 2018.

P. Papantoniou, E. Papadimitriou, and G. Yannis. Review of driving performance parameters critical for distracted driving research. *Transportation research procedia*, 25:1796–1805, 2017.

J. M. Pappas, S. R. Fishel, J. D. Moss, J. M. Hicks, and T. D. Leech. An eye-tracking approach to inattentional blindness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 1658–1662, Sage CA: Los Angeles, CA, 2005. SAGE Publications.

R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39 (2):230–253, 1997.

J. Park and W. Park. Functional requirements of automotive head-up displays: A systematic review of literature from 1994 to present. *Applied ergonomics*, 76:130–146, 2019.

H. Pashler. Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, 116(2):220, 1994.

H. Pashler and J. C. Johnston. Attentional limitations in dual-task performance. *Attention*, pages 155–189, 1998.

A. Pauzie. Head up display in automotive: a new reality for the driver. In *Design, User Experience, and Usability: Interactive Experience Design: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part III 4*, pages 505–516. Springer, 2015.

E. M. Peck, E. Carlin, and R. Jacob. Designing brain-computer interfaces for attention-aware systems. *Computer*, 48(10):34–42, 2015.

J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. Psychopy2: Experiments in behavior made easy. *Behavior research methods*, 51:195–203, 2019.

K. A. Pettijohn, C. Peltier, J. R. Lukos, J. N. Norris, and A. T. Biggs. Virtual and augmented reality in a simulated naval engagement: Preliminary comparisons of simulator sickness and human performance. *Applied ergonomics*, 89:103200, 2020.

L. Pipkorn, M. Dozza, and E. Tivesten. Driver visual attention before and after take-over requests during automated driving on public roads. *Human factors*, page 00187208221093863, 2022.

K. M. T. Pöhlmann, M. S. K. Auf Der Heyde, G. Li, F. Verstraten, S. A. Brewster, and M. McGill. Can visual motion presented in a vr headset reduce motion sickness for vehicle passengers? In *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 114–118, 2022a.

K. M. T. Pöhlmann, J. Föcker, P. Dickinson, A. Parke, and L. O'Hare. The relationship between vection, cybersickness and head movements elicited by illusory motion in virtual reality. *Displays*, 71:102111, 2022b.

K. M. T. Pöhlmann, H. A. Maior, J. Föcker, L. O'Hare, A. Parke, A. Ladowska, and P. Dickinson. I think i don't feel sick: Exploring the relationship between cognitive demand and cybersickness in virtual reality using fnirs. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2023.

K. M. T. Pöhlmann, G. Wilson, G. Li, M. Mcgill, and S. A. Brewster. From slow-mo to ludicrous speed: Comfortably manipulating the perception of linear in-car vr motion through vehicular translational gain and attenuation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2024.

I. Politis, S. Brewster, and F. Pollick. Evaluating multimodal driver displays of varying urgency. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 92–99, 2013.

I. Politis, S. Brewster, and F. Pollick. Speech tactons improve speech warnings for drivers. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 1–8, 2014a.

I. Politis, S. A. Brewster, and F. Pollick. Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 4067–4076, 2014b.

I. Politis, S. Brewster, and F. Pollick. To beep or not to beep? comparing abstract versus language-based multimodal driver displays. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3971–3980, 2015.

I. Politis, S. Brewster, and F. Pollick. Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 9(3):1–16, 2017.

A. K. Pradhan and D. Crundall. Hazard avoidance in young novice drivers: Definitions and a framework. In *Handbook of teen and novice drivers*, pages 81–94. CRC Press, 2016.

K. Pravossoudovitch, F. Cury, S. G. Young, and A. J. Elliot. Is red the colour of danger? testing an implicit red–danger association. *Ergonomics*, 57(4):503–510, 2014.

L. J. Prinzel III and M. Risser. Head-up displays and attention capture. Technical report, NASA, 2004.

Prolific. Prolific: Online participant recruitment for surveys and behavioral experiments. `https://www.prolific.co/`, 2024. Accessed: 2024-09-13.

Qualtrics. Qualtrics. `https://www.qualtrics.com/`, 2024. URL `https://www.qualtrics.com/`. Accessed: 2024-08-16.

RAC. Stopping distances made simple, 2023. URL `https://www.rac.co.uk/drive/advice/learning-to-drive/stopping-distances`.

J. Radlmayr, C. Gold, L. Lorenz, M. Farid, and K. Bengler. 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 annual meeting*, volume 58, pages 2063–2067. Sage Publications Sage CA: Los Angeles, CA, 2014.

J. Radlmayr, K. Brüch, K. Schmidt, C. Solbeck, and T. Wehner. Peripheral monitoring of traffic in conditionally automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 1828–1832. SAGE Publications Sage CA: Los Angeles, CA, 2018.

S. Rangelova and E. Andre. A Survey on Simulation Sickness in Driving Applications with Virtual Reality Head-Mounted Displays. *Presence: Teleoperators and Virtual Environments*, 27(1):15–31, 02 2018. doi: 10.1162/pres_a_00318. URL `https://doi.org/10.1162/pres_a_00318`.

J. Reason, A. Manstead, S. Stradling, J. Baxter, and K. Campbell. Errors and violations on the roads: a real distinction? *Ergonomics*, 33(10-11):1315–1332, 1990.

M. Resnick and W. Albert. The impact of advertising location and user task on the emergence of banner ad blindness: An eye-tracking study. *International Journal of Human-Computer Interaction*, 30(3): 206–219, 2014.

M. Rheu, J. Y. Shin, W. Peng, and J. Huh-Yoo. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, 37(1):81–96, 2021.

A. Riegler, A. Riener, and C. Holzmann. A systematic review of augmented reality applications for automated driving: 2009–2020. *PRESENCE: Virtual and Augmented Reality*, 28:87–126, 2019a.

A. Riegler, P. Wintersberger, A. Riener, and C. Holzmann. Augmented reality windshield displays and their potential to enhance user experience in automated driving. *i-com*, 18(2):127–149, 2019b.

A. Riegler, A. Riener, and C. Holzmann. A research agenda for mixed reality in automated vehicles. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*, pages 119–131, 2020.

A. Riegler, A. Riener, and C. Holzmann. Towards personalized 3d augmented reality windshield displays in the context of automated driving. *Frontiers in future transportation*, 3:1, 2022.

A. Robertson. Distance technologies car hud prototype: Hands-on at awe. *The Verge*, July 2024. URL `https://www.theverge.com/24185134/distance-technologies-car-hud-prototype-awe-hands-on`.

F. Roider, S. Rümelin, B. Pfleging, and T. Gross. The effects of situational demands on gaze, speech and gesture input in the vehicle. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 94–102, 2017.

P. Rubio-Fernandez, V. Shukla, V. Bhatia, S. Ben-Ami, and P. Sinha. Head turning is an effective cue for gaze following: Evidence from newly sighted individuals, school children and adults. *Neuropsychologia*, 174:108330, 2022.

K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer graphics forum*, volume 34, pages 299–326. Wiley Online Library, 2015.

M. L. Rusch, M. C. Schall Jr, P. Gavin, J. D. Lee, J. D. Dawson, S. Vecera, and M. Rizzo. Directing driver attention with augmented reality cues. *Transportation research part F: traffic psychology and behaviour*, 16:127–137, 2013.

E. Šabić, D. Henning, and J. MacDonald. Adaptive auditory alerts for smart in-vehicle interfaces. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 1545–1549. SAGE Publications Sage CA: Los Angeles, CA, 2019.

F. Sagberg and T. Bjørnskau. Hazard perception and driving experience among novice drivers. *Accident Analysis & Prevention*, 38(2):407–414, 2006.

K. J. Salubre and D. Nathan-Roberts. Takeover request design in automated driving: a systematic review. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 868–872. SAGE Publications Sage CA: Los Angeles, CA, 2021.

D. Salvucci and J. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research & Applications Symposium 2000*, Palm Beach Gardens, FL, November 6–8 2000. Association for Computing Machinery, New York, NY, USA.

D. Saredakis, A. Szpak, B. Birckhead, H. A. Keage, A. Rizzo, and T. Loetscher. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Frontiers in human neuroscience*, 14:96, 2020.

C. Schartmüller, K. Weigl, A. Löcken, P. Wintersberger, M. Steinhauser, and A. Riener. Displays for productive non-driving related tasks: Visual behavior and its impact in conditionally automated driving. *Multimodal Technologies and Interaction*, 5(4):21, 2021.

N. Schömig and B. Metz. Three levels of situation awareness in driving with secondary tasks. *Safety Science*, 56:44–51, 2013.

N. Schömig, K. Wiedemann, F. Naujoks, A. Neukum, B. Leuchtenberg, and T. Vöhringer-Kuhnt. An augmented reality display for conditionally automated driving. In *Adjunct proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, pages 137–141, 2018.

R. Schroeter and F. Steinberger. Pokémon drive: towards increased situational awareness in semi-automated driving. In *Proceedings of the 28th australian conference on computer-human interaction*, pages 25–29, 2016.

R. Schroeter, J. Oxtoby, and D. Johnson. Ar and gamification concepts to reduce driver boredom and risk taking behaviours. In *Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications*, pages 1–8, 2014.

E. H. Schumacher, T. L. Seymour, J. M. Glass, D. E. Fencsik, E. J. Lauber, D. E. Kieras, and D. E. Meyer. Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological science*, 12(2):101–108, 2001.

M. Schuß, L. Pizzoni, and A. Riener. Human or robot? exploring different avatar appearances to increase perceived security in shared automated vehicles. *Journal on Multimodal User Interfaces*, pages 1–20, 2024.

E. Schwalb. Analysis of hazards for autonomous driving. *Journal of Autonomous Vehicles and Systems*, 1(2):021003, 2021.

K. A. Schwarz, R. Pfister, M. Kluge, L. Weller, and W. Kunde. Do we see it or not? sensory attenuation in the visual domain. *Journal of Experimental Psychology: General*, 147(3):418, 2018.

Y. Sela and Y. Amichai-Hamburger. “baby, i can’t drive my car”: How controllability mediates the relationship between personality and the acceptance of autonomous vehicles? *International Journal of Human–Computer Interaction*, pages 1–11, 2023.

D. Shapiro. Meet nvidia drive concierge and drive chauffeur, 2021. URL `https://blogs.nvidia.com/blog/drive-chauffeur-and-concierge/`. Accessed: 2024-07-10.

S. Sheikhi and J.-M. Odobez. Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions. *Pattern Recognition Letters*, 66:81–90, 2015.

D. J. Simons. Attentional capture and inattentional blindness. *Trends in cognitive sciences*, 4(4):147–155, 2000.

D. J. Simons. Inattentional blindness. *Scholarpedia*, 2(5):3244, 2007.

D. J. Simons and C. F. Chabris. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28(9):1059–1074, 1999. ISSN 03010066. doi: 10.1068/p281059.

C. Smith. Bejeweled player count and stats (2023), 2023. URL `https://videogamesstats.com/bejeweled/`.

M. Smith, J. Streeter, G. Burnett, and J. L. Gabbard. Visual search tasks: the effects of head-up displays on driving and task performance. In *Proceedings of the 7th international conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 80–87, 2015.

M. Smith, J. L. Gabbard, and C. Conley. Head-up vs. head-down displays: examining traditional methods of display assessment while driving. In *Proceedings of the 8th international conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 185–192, 2016.

M. Smith, K. Bagalkotkar, J. L. Gabbard, D. R. Large, and G. Burnett. Isolating the effect of off-road glance duration on driving performance: An exemplar study comparing hdd and hud in different driving scenarios. *Human factors*, 65(5):833–845, 2023.

S. Soares, A. Lobo, S. Ferreira, L. Cunha, and A. Couto. Takeover performance evaluation using driving simulation: a systematic review and meta-analysis. *European transport research review*, 13:1–18, 2021.

Society of Automotive Engineers. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *SAE International*, 4970(724):1–5, 2018.

M. Sodhi, B. Reimer, and I. Llamazares. Glance analysis of driver eye movements to evaluate distraction. *Behavior Research Methods, Instruments, & Computers*, 34:529–538, 2002.

R. Srour Zreik, M. Harvey, and S. A. Brewster. Age matters: Investigating older drivers’ perception of level 3 autonomous cars as a heterogeneous age group. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 13–18, 2023.

R. Srour Zreik, T. Goodge, M. Harvey, and S. Brewster. Where's my tor?: Evaluating the effect of take-over request source on older drivers' control transition in level 3 cars. *Proceedings of the ACM on Human-Computer Interaction*, 8(MHCI):1–25, 2024.

K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, volume 41, pages 1138–1142. SAGE Publications Sage CA: Los Angeles, CA, 1997.

J. K. Stefanucci, D. Brickler, H. C. Finney, E. Wilson, T. Drew, and S. H. Creem-Regehr. Effects of simulated augmented reality cueing in a virtual navigation task. *Frontiers in Virtual Reality*, 3:971310, 2022.

F. Steinberger, R. Schroeter, and C. N. Watling. From road distraction to safe driving: Evaluating the effects of boredom and gamification on driving behaviour, physiological arousal, and subjective experience. *Computers in Human Behavior*, 75:714–726, 2017.

D. L. Strayer, F. A. Drews, and W. A. Johnston. Cell phone-induced failures of visual attention during simulated driving. *Journal of experimental psychology: Applied*, 9(1):23, 2003.

T. Strobach. The dual-task practice advantage: Empirical evidence and cognitive mechanisms. *Psychonomic bulletin & review*, 27(1):3–14, 2020.

H. Summala, E. Pasanen, M. Räsänen, and J. Sievänen. Bicycle accidents and drivers' visual search at left and right turns. *Accident Analysis & Prevention*, 28(2):147–153, 1996.

N. Sun and J. Botev. Intelligent autonomous agents and trust in virtual reality. *Computers in Human Behavior Reports*, 4:100146, 2021.

S. Tamura, N. Ohshima, K. Hasegawa, and M. Okada. Design and evaluation of attention guidance through eye gazing of "namida" driving agent. *Journal of Robotics and Mechatronics*, 33(1):24–32, 2021.

Q. Tang and G. Guo. Is the eye movement pattern the same? the difference between automated driving and manual driving. In *Human Aspects of IT for the Aged Population. Social Media, Games and Assistive Environments: 5th International Conference, ITAP 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part II*, volume 21, pages 554–563. Springer International Publishing, 2019.

TechCrunch. Waymo expands robotaxi coverage in los angeles and san francisco. `https://techcrunch.com/2024/08/06/waymo-expands-robotaxi-coverage-in-los-angeles-and-san-francisco/`, 2024. Accessed: 2024-08-07.

A. Telpaz, B. Rhindress, I. Zelman, and O. Tsimhoni. Haptic seat for automated driving: preparing the driver to take control effectively. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 23–30, 2015.

O. Tengilimoglu, O. Carsten, and Z. Wadud. Implications of automated vehicles for physical road environment: A comprehensive review. *Transportation research part E: logistics and transportation review*, 169:102989, 2023.

Tesla. Touchscreen | meet your model 3, 2023. URL https://www.tesla.com/support/videos/watch/touchscreen-meet-your-model-3.

The European Parliament. Commission implementing regulation (eu) 2022/1426 of 5 august 2022 laying down rules for the application of regulation (eu) 2019/2144 of the european parliament and of the council as regards uniform procedures and technical specifications for the type-approval of the automated driving system (ads) of fully automated vehicles (text with eea relevance), August 2022. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R1426.

P. Thiffault and J. Bergeron. Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, 35(3):381–391, 2003.

M. A. Thornton and A. R. Conway. Working memory for social information: Chunking or domain-specific buffer? *Neuroimage*, 70:233–239, 2013.

P.-H. Ting, J.-R. Hwang, J.-L. Doong, and M.-C. Jeng. Driver fatigue and highway driving: A simulator study. *Physiology & behavior*, 94(3):448–453, 2008.

M. Tisshaw. Augmented reality heads-up displays for cars are finally a real thing. *Ars Technica*, July 2020. URL https://arstechnica.com/cars/2020/07/augmented-reality-heads-up-displays-for-cars-are-finally-a-real-thing/.

H. Togwell, M. McGill, G. Wilson, D. Medeiros, and S. A. Brewster. In-car gaming: Exploring the use of ar headsets to leverage passenger travel environments for mixed reality gameplay. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, 2022.

J. E. Törnros and A. K. Bolling. Mobile phone use—effects of handheld and handsfree phones on driving performance. *Accident Analysis & Prevention*, 37(5):902–909, 2005.

A. Treisman and A. Davies. Divided attention to ear and eye. *From perception to consciousness: Searching with Anne Treisman*, 24, 2012.

S. Trösterer, M. Gärtner, M. Wuchse, B. Maurer, A. Baumgartner, A. Meschtscherjakov, and M. Tscheligi. Four eyes see more than two: Shared gaze in the car. In *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15*, pages 331–348. Springer, 2015.

S. Trösterer, B. Streitwieser, A. Meschtscherjakov, and M. Tscheligi. Shared gaze while driving: how drivers can be supported by an led-visualization of the front-seat passenger's gaze. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part II 17*, pages 329–350. Springer, 2019.

I. P. Tussyadiah, F. J. Zach, and J. Wang. Do travelers trust intelligent service robots? *Annals of Tourism Research*, 81:102886, 2020.

UK and Scottish Law Commissions. Automated vehicles: Consultation paper 3-a regulatory framework for automated vehicles a joint consultation paper, 2020, December 2020. URL https://www.scotlawcom.gov.uk/files/7516/6902/3022/CP3_Upload_final_17_Dec_2020.pdf.

UK Government. Statistical data set: Reported road collisions, vehicles and casualties tables for great britain, December 2022a. URL https://www.gov.uk/government/statistical-data-sets/reported-road-accidents-vehicles-and-casualties-tables-for-great-britain.

UK Government. The road vehicles (construction and use) (amendment) regulations 2022. https://www.legislation.gov.uk/en/uksi/2022/81/made, 2022b. Accessed: 2024-08-07.

UK Government. Self-driving vehicles set to be on roads by 2026 as automated vehicles act becomes law, 2024. URL https://www.gov.uk/government/news/self-driving-vehicles-set-to-be-on-roads-by-2026-as-automated-vehicles-act-becomes-law. Accessed: 2024-09-13.

UK Government Statistics. Road safety factors: Initial analysis, 2024. URL https://www.gov.uk/government/statistics/road-safety-factors-initial-analysis/road-safety-factors-initial-analysis. Accessed: 2024-08-16.

UK Parliament. Autonomous vehicle bill 2023, 2023. URL https://bills.parliament.uk/publications/52908/documents/3984. 24/11/2023.

UK Parliament. Automated vehicles act 2024, May 2024. URL https://bills.parliament.uk/bills/3506.

G. Underwood. Visual attention and the transition from novice to advanced driver. *Ergonomics*, 50(8): 1235–1249, 2007.

G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, 46(6):629–646, 2003.

U.S. Department of Transportation. Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0, 2021. URL https://www.transportation.gov/sites/dot.gov/files/2021-01/USDOT_AVCP.pdf. Accessed: 2024-09-13.

T. Van Veen, J. Karjanto, and J. Terken. Situation awareness in automated vehicles through proximal peripheral light signals. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*, pages 287–292, 2017.

B. M. Velichkovsky, S. M. Dornhoefer, S. Pannasch, and P. J. Unema. Visual fixations and level of attentional processing. In *Proceedings of the 2000 symposium on eye tracking research & applications*, pages 79–85, 2000.

P. Ventsislavova and D. Crundall. The hazard prediction test: A comparison of free-response and multiple-choice formats. *Safety Science*, 2018. ISSN 18791042. doi: 10.1016/j.ssci.2018.06.004.

P. Ventsislavova, D. Crundall, T. Baguley, C. Castro, A. Gugliotta, P. Garcia-Fernandez, W. Zhang, Y. Ba, and Q. Li. A comparison of hazard perception and hazard prediction tests across china, spain and the uk. *Accident Analysis & Prevention*, 122:268–286, 2019.

P. Ventsislavova Petrova. *Development of a cross-cultural hazard perception test*. Nottingham Trent University (United Kingdom), 2019.

F. M. Verberne, J. Ham, and C. J. Midden. Trusting a virtual driver that looks, acts, and thinks like you. *Human factors*, 57(5):895–909, 2015.

F. Vicente, Z. Huang, X. Xiong, F. De la Torre, W. Zhang, and D. Levi. Driver gaze tracking and eyes off the road detection system. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):2014–2027, 2015.

M. Vollrath, T. Meilinger, and H.-P. Krüger. How the presence of passengers influences the risk of a collision with another vehicle. *Accident Analysis & Prevention*, 34(5):649–654, 2002.

M. Walch, K. Mühl, J. Kraus, T. Stoll, M. Baumann, and M. Weber. From car-driver-handovers to cooperative interfaces: Visions for driver–vehicle interaction in automated driving. *Automotive user interfaces: Creating interactive experiences in the car*, pages 273–294, 2017.

M. Walch, D. Lehr, M. Colley, and M. Weber. Don't you see them? towards gaze-based interaction adaptation for driver-vehicle cooperation. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: Adjunct proceedings*, pages 232–237, 2019.

G. H. Walker, N. A. Stanton, T. A. Kazi, P. M. Salmon, and D. P. Jenkins. Does advanced driver training improve situational awareness? *Applied ergonomics*, 40(4):678–687, 2009.

H. E. Walker, R. A. Eng, and L. M. Trick. Dual-task decrements in driving performance: The impact of task type, working memory, and the frequency of task performance. *Transportation research part F: traffic psychology and behaviour*, 79:185–204, 2021.

C. D. Wallbridge, V. Marcinkiewicz, Q. Zhang, and P. Morgan. Towards anthropomorphising autonomous vehicles: speech and embodiment on trust and blame after an accident, 2022.

B. Wandtner, N. Schömig, and G. Schmidt. Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human factors*, 60(6):870–881, 2018.

C. Wang, S. Steeghs, D. Chakraborty, A. Gorle, D. Dey, S. van de Star, A. Sudhakaran, J. Terken, and J. Hu. Designing for enhancing situational awareness of semi-autonomous driving vehicles. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*, pages 228–229, 2017a.

D. Wang, C. Lu, M. Zeng, D. Luo, and M. Wang. Can speeding behavior be reduced? the effect of the display mode and color change of ui. In *International Conference on Computer Application and Information Security (ICCAIS 2022)*, volume 12609, pages 390–398. SPIE, 2023.

M. Wang, S. Lundgren Lyckvi, C. Chen, P. Dahlstedt, and F. Chen. Using advisory 3d sound cues to improve drivers' performance and situation awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2814–2825, 2017b.

M. Wang, S. C. Lee, G. Montavon, J. Qin, and M. Jeon. Conversational voice agents are preferred and lead to better driving performance in conditionally automated vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 86–95, 2022a.

Y. Wang, Y. Wu, C. Chen, B. Wu, S. Ma, D. Wang, H. Li, and Z. Yang. Inattentional blindness in augmented reality head-up display-assisted driving. *International Journal of Human–Computer Interaction*, 38(9):837–850, 2022b.

M. D. Ward and W. S. Helton. Dual-task interference while receiving information on a head mounted display and manual tracking with and without auditory warnings. *Applied ergonomics*, 101:103713, 2022.

A. C. Warden, C. D. Wickens, D. Mifsud, S. Ourada, B. A. Clegg, and F. R. Ortega. Visual search in augmented reality: Effect of target cue type and location. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 66, pages 373–377. Sage CA: Los Angeles, CA: SAGE Publications, 2022.

B. W. Weaver and P. R. DeLucia. A systematic review and meta-analysis of takeover performance during conditionally automated driving. *Human factors*, 64(7):1227–1260, 2022.

C. D. Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.

J. Wiemer, A. B. Gerdes, and P. Pauli. The effects of an unexpected spider stimulus on skin conductance responses and eye movements: an inattentional blindness study. *Psychological research*, 77:155–166, 2013.

C. Wilson, D. Gyi, A. Morris, R. Bateman, and H. Tanaka. Non-driving related tasks and journey types for future autonomous vehicle owners. *Transportation research part F: traffic psychology and behaviour*, 85:150–160, 2022.

P. Wintersberger, A.-K. Frison, A. Riener, and T. v. Sawitzky. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 27(1):46–62, 02 2018a. doi: 10.1162/pres_a_00320. URL `https://doi.org/10.1162/pres_a_00320`.

P. Wintersberger, A. Riener, C. Schartmüller, A.-K. Frison, and K. Weigl. Let me finish before i take over: Towards attention aware device integration in highly automated vehicles. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, pages 53–65, 2018b.

J. M. Wolfe, A. Kosovicheva, and B. Wolfe. Normal blindness: when we look but fail to see. *Trends in Cognitive Sciences*, 2022.

J. Wolffsohn, N. McBrien, G. K. Edgar, and T. Stout. The influence of cognition and age on accommodation, detection rate and response times when using a car head-up display (hud). *Ophthalmic and physiological optics*, 18(3):243–253, 1998.

C. Wu, H. Wu, N. Lyu, and M. Zheng. Take-over performance and safety analysis under different scenarios and secondary tasks in conditionally automated driving. *IEEE Access*, 7:136924–136933, 2019.

H. Wu, C. Wu, N. Lyu, and J. Li. Does a faster takeover necessarily mean it is better? a study on the influence of urgency and takeover-request lead time on takeover performance and safety. *Accident analysis & prevention*, 171:106647, 2022.

Y. Wu, K. Kihara, Y. Takeda, T. Sato, M. Akamatsu, S. Kitazaki, K. Nakagawa, K. Yamada, H. Oka, and S. Kameyama. Eye movements predict driver reaction time to takeover request in automated driving: A real-vehicle study. *Transportation research part F: traffic psychology and behaviour*, 81:355–363, 2021.

Y. Wu, L. Ye, X. Zheng, H. Zhu, and W. Xiang. Dangerous slime: A game for improving situation awareness in automated driving. In *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 136–144, 2023a.

Z. Wu, L. Zhao, G. Liu, J. Chai, J. Huang, and X. Ai. The effect of ar-hud takeover assistance types on driver situation awareness in highly automated driving: A 360-degree panorama experiment. *International Journal of Human–Computer Interaction*, pages 1–18, 2023b.

K. Yamaguchi, M. Nakayama, Q. Sun, and J. Xia. Prediction of cognitive performance of drivers using eye fixation behaviours. In *2019 23rd International Conference Information Visualisation (IV)*, pages 186–190. IEEE, 2019.

A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 131–140, 2008.

J. Yang, C. Liu, P. Chu, X. Wen, and Y. Zhang. Exploration of the relationships between hazard perception and eye movement for young drivers. *Journal of advanced transportation*, 2021(1):6642999, 2021a.

J. H. Yang, S. C. Lee, C. Nadri, J. Kim, J. Shin, and M. Jeon. Multimodal displays for takeover requests. *User Experience Design in the Era of Automated Driving*, pages 397–424, 2022.

Y. Yang, B. Karakaya, G. C. Dominioni, K. Kawabe, and K. Bengler. An hmi concept to improve driver's visual behavior and situation awareness in automated vehicle. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 650–655. IEEE, 2018.

Z. Yang, Q. Yu, W. Zhang, and H. Shen. A comparison of experienced and novice drivers' rear-end collision avoidance maneuvers under urgent decelerating events. *Transportation research part F: traffic psychology and behaviour*, 76:353–368, 2021b.

G. Yannis, E. Papadimitriou, and P. Papantoniou. Distracted driving and mobile phone use: Overview of impacts and countermeasures. *Commun. Technol. Road Saf*, pages 1–23, 2014.

M. Yeh and C. D. Wickens. Target cueing in augmented reality: A comparison of head mounted with hand-held displays. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 43, pages 1228–1232. Sage CA: Los Angeles, CA: SAGE Publications, 1999.

S. Yiğit-Elliott, J. Palmer, and C. M. Moore. Distinguishing blocking from attenuation in visual selective attention. *Psychological Science*, 22(6):771–780, 2011.

S. H. Yoon and Y. G. Ji. Non-driving-related tasks, workload, and takeover performance in highly automated driving contexts. *Transportation research part F: traffic psychology and behaviour*, 60:620–631, 2019.

A. H. Young, A. K. Mackenzie, R. L. Davies, and D. Crundall. Familiarity breeds contempt for the road ahead: The real-world effects of route repetition on visual attention in an expert driver. *Transportation research part F: traffic psychology and behaviour*, 57:4–9, 2018.

K. Young, M. Regan, M. Hammer, et al. Driver distraction: A review of the literature. *Distracted driving*, 2007:379–405, 2007.

N. Zangi, R. Srour-Zreik, D. Ridel, H. Chassidim, and A. Borowsky. Driver distraction and its effects on partially automated driving performance: A driving simulator study among young-experienced drivers. *Accident Analysis & Prevention*, 166:106565, 2022.

K. Zeeb, A. Buchner, and M. Schrauf. What determines the take-over time? an integrated model approach of driver take-over after automated driving. *Accident analysis & prevention*, 78:212–221, 2015.

K. Zeeb, A. Buchner, and M. Schrauf. Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident analysis & prevention*, 92:230–239, 2016.

B. Zhang, J. De Winter, S. Varotto, R. Happee, and M. Martens. Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation research part F: traffic psychology and behaviour*, 64:285–307, 2019.

Y. Zhang, T. Yang, X. Zhang, Y. Zhang, and Y. Sun. Effects of full windshield head-up display on visual attention allocation. *Ergonomics*, 64(10):1310–1321, 2021.

Appendix A

List of Hazard Prediction Clips

Table A.1: List of each Hazard Clip, with the foil answers, the correct answer in bold, the length of each clip and the screen location of where the hazard starts to become visible.

Clip	Length	Corr	Answer1	Answer2	Answer3	Answer4	Location (x,y)
Clip01WHN	35.03	4	The cyclist on the left pulls into your lane	A pedestrian steps out from behind the van parked on the left	A pedestrian steps out from between cars in the oncoming lane	**A car pulls out from the side road on the right**	Centre (1040,820)
Clip02WHN	30.05	2	The blue car on the left pulls off in front of you	**Pedestrians step out into the road from the right**	The white lorry on the right pulls round into your lane	A cyclist pulls out from behind the parked cars on the left	Centre Right (1145,830)
Clip03WHN	22.72	1	**The door of the car parked on the left swings open and the driver steps out**	A pedestrian at the bus stop steps out into the road	A car pulls out from the side road on the right	An oncoming car comes round the corner in your lane	Centre Left (830,730)
Clip04WHN	22.93	3	A pedestrian steps out from behind the car parked on the left	A pedestrian steps out from behind the van parked on the right	**A van pulls out from the side road on the left**	A child runs into the road from the school on the right	Centre Left (900,840)
Clip05WHN	28.13	1	**A pedestrian steps out from behind the car parked on the left**	A pedestrian walks across the road from the right	A car parked on the left pulls off in front of you	A cyclist pulls into your lane from the right	Centre (800,1000)
Clip06WHN	28.05	2	The white parked car on the left pulls off in front of you	**The white van pulls out of the side road on the left**	The black oncoming car pulls across your lane into the side road on the left	A delivery worker steps out from behind the white van parked on the left	Centre (970,800)
Clip07WHN	36.25	4	The black parked car on the left pulls off in front of you	A pedestrian steps into the road from the right	An car takes a wrong turn and pulls into your lane	**A pedestrian steps into the road from behind the car parked on the left**	Centre Left (590,750)
Clip08WHN	24.08	3	The black parked car on the left pulls off in front of you	The pedestrians in the bus stop step into the road from the right	**A car stopped in the road ahead forces you to swerve to avoid it**	The cyclist on the left swerves into the road	Centre (800,1030)
Clip09WHN	15.35	4	A pedestrian steps out from behind the car parked on the right	The red car parked on the left pulls off in front of you	The white van on the right speeds up and pulls in front of yo	**A pedestrian steps into the road from behind the car parked on the left**	Centre (1010,810)

Clip	Length	Corr	Answer1	Answer2	Answer3	Answer4	Location (x,y)
Clip10WHN	29.23	1	**A car door opens from the left and the driver steps out**	A car pulls out from the road on the right	The blue car parked on the left pulls off in front of you	A delivery van reverses into the road from the right	Centre Left (930,810)
Clip11WHN	42.42	3	A delivery worker steps into the road from behind the white van on the left	The pedestrian on the right runs across the road	**The car in front brakes to allow a parked car to pull off**	A carpet tube falls into the road from the delivery van on the left	Centre (1045,790)
Clip12WHN	23.27	3	A pedestrian steps out from the left at the crossing	A car pulls out from the road on the left	**The oncoming car pulls across the road in front of you**	The red parked car pulls off and performs a u-turn in front of you	Centre (1020,785)
Clip13WHN	45.77	4	A van pulls out into the road from the side road on the left	A pedestrian steps out from behind the car parked on your left	The grey car on the left pulls off in front of you	**The car ahead brakes in the road, forcing you to slow down to avoid it**	Centre (950,780)
Clip14WHN	35.98	2	A police car pulls out into the road from the side road on the left	**A pedestrian walks across the road from the right**	The white car on the left pulls off in front of you	A group of school children run across the road from the left	Centre Right (1110,750)
Clip15WHN	52.2	2	A delivery worker steps into the road from behind the white van on the left	**Oncoming cyclists encroach on your lane**	A pedestrian steps out from behind the car parked on the right	The door of the white car parked on the right opens and the driver steps out	Centre Right (1065,765)
Clip16WHN	26.15	1	**An oncoming van comes round the corner in your lane**	A car pulls out into the road from the side road on the left	The pedestrians in the bus stop step into the road from the left	A pedestrian steps out from between cars in the oncoming lane	Centre Right (1284,775)
Clip17WHN	19.93	4	The car in the right lane pulls sharply across into your lane	A pedestrian at the bus on the left top steps out into the road	The white van on the left pulls off in front of you	**A delivery worker steps into the road from behind the white van on the left**	Centre (940,765)
Clip18WHN	34.73	3	A pedestrian steps out from the left at the bus stop	The silver car parked on the left pulls off in front of you	**The black van ahead brakes in the road, forcing you to slow down to avoid it**	A delivery van reverses into the road from the right	Centre (1115,740)
Clip19WHN	41.27	2	A car jumps the lights and pulls across you from the right	**Pedestrians step out into the road from the left at the crossing**	A cyclist swerves into the road from the left	The white car parked on the left pulls off in front of you	Centre (950,770)

Clip	Length	Corr	Answer1	Answer2	Answer3	Answer4	Location (x,y)
Clip20WHN	29.65	3	A pedestrian steps out from behind the car parked on the left	A car pulls out of the side road on the left	**The white car on the right reverses into the road**	A pedestrian steps out from behind the car parked on the right	Centre Right (1225,770)
Clip21WHN	14.95	4	The white van on the left starts to reverse into the road	Pedestrians step out from the left	The silver van on the right pulls out in front of you	**The white car ahead brakes, forcing you to go around it**	Centre (930,730)
Clip22WHN	21.48	2	The blue car on the garage forecourt pulls out in front of you	**The white car on the left pulls into your lane**	The car ahead brakes sharply	A pedestrian steps onto the road from the central reservation on the right	Centre Left (830,810)
Clip23WHN	26.63	3	The white car on the left pulls off in front of you	A worker steps into the road from the left	**A white van pulls out from the road on the right**	The blue car ahead brake suddenly for congestion	Centre Right (1280,775)
Clip24WHN	39.5	2	The oncoming car pulls across your lane	**A van stopped in the road ahead blocks your path**	The black car pulls out from the side road on the left	The jogger runs across the road in front of you from the left	Centre (970,720)
Clip25WHN	41.15	1	**A lorry stopped in the road ahead blocks your path**	The silver car on the left pulls off in front of you	A lorry pulls out from the recycling centre ahead on the left	The black car ahead starts reversing towards you	Centre Right (1340,725)
Clip26WHN	36.08	3	The red car on the left starts to pull out in front of you	The white van on the right pulls across into your lane	**The blue car on the right pulls across into your lane**	A car pulls out from in front of the cars parked on the left	Centre Right (1155,790)
Clip27WHN	55.08	1	**A car pulls out from in front of the cars parked on the left**	Pedestrians step into the road from the left	The red car parked across the street performs a u-turn in front of you	The cyclist on the right sets off into the road in front of you	Centre (910,775)
Clip28WHN	38.82	4	The oncoming car pulls across your lane	The grey car on the left starts to pull out in front of you	A car pulls out from in front of the cars parked on the left	**A pedestrian steps out from behind the car parked on the left**	Centre (1010,765)
Clip29WHN	40.85	4	A pedestrian steps out from behind the car parked on the left	A pedestrian steps out from the right	The red car on the left pulls off in front of you	**The black car ahead starts reversing towards you**	Centre (940,780)
Clip30WHN	27.4	1	**A car pulls out from the road on the right**	The blue car on the left pulls off in front of you	A cyclist pulls out from the side road on the left	A pedestrian steps out from the right	Centre Right (1125,785)

Clip	Length	Corr	Answer1	Answer2	Answer3	Answer4	Location (x,y)
Clip31WHN	39.68	2	The taxi on the left pulls off in front of you	**Pedestrians step into the road from the right**	The van ahead brakes suddenly	A pedestrian steps out from the left	Centre (1010,740)
Clip32WHN	19.82	1	**A car pulls out from the right behind the oncoming bus**	The police van on the left turns it's sirens on and pulls off in front of you	The bus pulls out into your lane to go around the stationary bus	A pedestrian steps out into the road from the right behind the bus	Centre (1020,745)
Clip33WHN	31.33	2	The red car on the left starts to pull out in front of you	**The oncoming car encroaches on your lane**	A pedestrian steps out into the road from the right behind the bus	A car reverses out of the driveway on the left	Centre (990,765)
Clip34WHN	60.36	3	A car pulls out from the side road on the left	The white car on the left pulls into your lane	**A cyclist pulls out from the from the central reservation on the right**	A car door opens from the parked car on the left	Centre (1000,735)
Clip35WHN	59.22	4	The car ahead indicating to turn left suddenly swings back out infront of you	The yellow car on the left starts reversing into the road	The car ahead brakes sharply to stop for the bus	**A pedestrian runs across the road from the right**	Centre Right (1120,740)
Clip36WHN	31.47	1	**A car reverses backwards into the road ahead of you**	A pedestrian steps out from the right	A pedestrian steps out from behind the parked cars on the left	The black car parked on the left pulls off in front of you	Centre (995,850)
Clip37WHN	40.73	3	A car pulls out from the side road on the left	A pedestrian steps out from the left	**A group of school children run into the road from the right**	The silver car parked on the left pulls off in front of you	Centre (1070,800)
Clip38WHN	33.88	4	The white van pulls out of the side road on the left	The oncoming car encroaches on your lane	A pedestrian steps into the road from the bus stop on the left	**The white car pulls out of the road on the right**	Centre Right (1150,855)
Clip39WHN	41.77	2	The red car on the left starts to pull out in front of you	**A car from the centre lane pulls into your lane**	The oncoming van encroaches on your lane	A pedestrian steps out from between the parked cars on the left	Centre (1090,805)
Clip40WHN	33.75	1	**The oncoming van encroaches on your lane**	The cyclist on the left pull out into the road	The white car parked ahead on the left pulls out in front of you	The red car parked on the right performs a u turn and pulls out in front of you	Centre (1100,850)

Appendix B

Hazard Clip Validation Coefficients

Clip Number	Experienced Score	Novice Score	Difference in Score	Prediction Coefficient
Clip40	0.98	0.82	0.16	0.0659
Clip18	0.98	0.92	0.07	0.0582
Clip38	0.52	0.16	0.36	0.0515
Clip02	0.98	0.85	0.13	0.046
Clip20	0.72	0.46	0.26	0.0446
Clip24	0.52	0.2	0.33	0.0441
Clip39	0.89	0.62	0.26	0.0415
Clip12	0.98	0.85	0.13	0.0404
Clip33	0.79	0.49	0.3	0.0399
Clip31	0.85	0.62	0.23	0.0388
Clip26	0.95	0.79	0.16	0.038
Clip13	0.95	0.79	0.16	0.037
Clip21	0.59	0.3	0.3	0.0366
Clip06	0.92	0.72	0.2	0.0357
Clip23	0.85	0.66	0.2	0.0355
Clip35	0.66	0.39	0.26	0.0322
Clip37	0.95	0.82	0.13	0.0316
Clip17	0.75	0.52	0.23	0.0281
Clip28	0.69	0.46	0.23	0.0257
Clip16	0.75	0.52	0.23	0.0251
Clip07	0.82	0.69	0.13	0.0246
Clip04	0.89	0.72	0.16	0.0214
Clip03	0.66	0.43	0.23	0.0203
Clip27	0.89	0.72	0.16	0.02
Clip32	0.56	0.43	0.13	0.0194
Clip29	0.92	0.79	0.13	0.0166
Clip01	0.82	0.69	0.13	0.0147
Clip15	0.79	0.69	0.1	0.0127
Clip11	0.72	0.62	0.1	0.0123
Clip08	0.59	0.49	0.1	0.0115
Clip19	0.95	0.89	0.07	0.0115
Clip14	0.92	0.82	0.1	0.0102
Clip10	0.85	0.75	0.1	0.0092
Clip25	0.72	0.62	0.1	0.0007
Clip05	0.82	0.75	0.07	-0.0008
Clip09	0.69	0.62	0.07	-0.0015
Clip22	0.69	0.62	0.07	-0.0074
Clip36	0.66	0.66	0	-0.0142
Clip34	0.33	0.49	-0.16	-0.0266
Clip30	0.89	0.92	-0.03	-0.0455

Table B.1: Clip Ratings: List of each Hazard Clip, the probability of each driver group of getting the answer correct and the prediction coefficients that describe the discriminatory power of each clip.

Appendix C

Experiment 3: Analysis Tables

Experiment 3 NASA TLX Ratings Analysis Table

NASA TLX Scale	Condition	Score		Comparison			
		Mean	SD	AR HUD	Cue AR HUD	AR HDD	Tablet HDD
	Baseline	78.08333	34.48871	**p = .26**	**p = .51**	**p< .001*** **	**p = .005** **
Total	AR HUD	106.5	41.21102	\	p = .51	p = .061	p = .33
(F(2.91, 64.02) =20.59, p < .001, η^2 = 0.31)	Cue AR HUD	87	34.10597		\	**p = .002** **	**p = .028* **
	AR HDD	140.25	37.8547			\	p = .51
	Tablet HDD	131.79167	37.57194				\
	Baseline	38.20833	21.71151	p = .22	p = .7	**p = .001** **	**p = .001** **
Mental Demand	AR HUD	54.04167	24.53477	\	p = .9	**p = .22**	p = .22
(F(2.9, 63.83) =16.91, p< .001, η^2 = 0.22)	Cue AR HUD	46.375	22.45539		\	p = .054	p = .054
	AR HDD	68.58333	21.86702			\	p = .9
	Tablet HDD	65.95833	21.94554				\
	Baseline	5.764706	7.258869	p = .29	p = .29	**p = .016* **	**p = .001** **
Physical Demand	AR HUD	15	14.990906	\	p = .87	p = .3	p = .3
(F(4, 56) = 7.32, p < .001, η^2 = 0.21)	Cue AR HUD	**13.285714**	**10.555297**		\	**p = .29**	p = .29
	AR HDD	28.5	25.761447			\	p = .87
	Tablet HDD	23.291667	19.501347				\
	Baseline	21.3913	13.93406	p = .02*	p = .3	p = .018*	p = .12
Temporal Demand	AR HUD	**41.17391**	**23.24113**	\	**p = .52**	**p = .52**	p = .6
(F(2.77, 58.22) = 11.28, p < .001, η^2 = 0.2)	Cue AR HUD	32	18.60108		\	p = .21	p = .52
	AR HDD	50.20833	27.15972			\	p = .6
	Tablet HDD	45.34783	27.01449				\
	Baseline	**39.875**	**23.10903**	**p = .37**	**p = .99**	**p < .001*** **	**p = .014* **
Overall Performance	AR HUD	52.45833	20.78666	\	p = .22	**p = .02* **	p = .22
(F(2.57, 56.43) = 13.31, p < .001, η^2 = 0.28)	Cue AR HUD	40.625	15.85755		\	**p< .001*** **	**p = .004** **
	AR HDD	71.66667	21.07062			\	p = .97
	Tablet HDD	**65.83333**	**24.83628**				\
	Baseline	36	24.0163	p = .16	p = .42	**p = .002** **	p = .09
Effort	AR HUD	53.125	27.30554	\	p = .42	p = .42	p = .81
(F(4, 88) = 13.71, p < .001, η^2 = 0.19)	Cue AR HUD	42.45833	20.08023		\	p = .02*	p = .42
	AR HDD	66.41667	25.02506			\	p = .42
	Tablet HDD	55.41667	23.72289				\
	Baseline	22.18182	23.64748	p = .44	p = .45	**p = .003** **	p = .11
	AR HUD	39.43478	29.63233	\	p = .45	p = .35	p = .44
Frustration	Cue AR HUD	31.3913	26.96587		\	**p = .048* **	p = .44
(F(2.79, 53.07) = 14.49, p < .001, η^2 = 0.23)	AR HDD	58.41667	33.23947			\	p = .45
	Tablet HDD	50.81818	35.03146				\

Table C.1: A table showing the full list of repeated measures ANOVA and pairwise comparisons for the NASA TLX ratings in Experiment 3.

Experiment 3 Simulator Sickness Tables

Timepoint	Mean Total SSQ Score	SD	Min	Max
Baseline	7.324167	9.644823	0	37.4
T1	7.0125	10.59627	0	48.62
T2	7.791667	9.221744	0	37.4
T3	9.038333	11.07909	0	44.88
T4	12.310833	13.88381	0	48.62
T5	13.401667	18.64844	0	86.02

Table C.2: A table showing a summary of Total SSQ Scores in Experiment 3

Model Intercept	Baseline			T1			T2			T3			T4			T5		
	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig
Baseline	7.32	2.5		-0.312	1.942	p = .872	0.292	2.248	p = .897	1.198	2.279	p = .6426	4.377	2.28	p = .06	5.617	**2.323**	**p = .017***
T1				7.013	2.498		0.604	2.248	p = .789	1.51	2.279	p = .509	4.688	**2.28**	**p = .042***	5.929	**2.323**	**p = .012***
T2							7.617	2.743		0.906	1.947	p = .643	4.084	**1.952**	**p = .039***	5.325	**1.972**	**p = .008****
T3										8.522	2.768		3.178	1.943	p = .105	4.419	**1.957**	**p = .026***
T4													11.701	2.77		1.241	1.96	p = .528

Table C.3: A table showing the Model Estimates, Standard Error (SE) and p values obtained through Wald's approximation for each of the LME models of Total SSQ scores at each timepoint of Experiment 3.

Appendix D

Experiment 4: Analysis Tables

Experiment 4 NASA TLX Ratings Analysis Table

NASA TLX Scale	Condition	Score		Comparison		
		Mean	SD	AR HUD	Cue AR HUD	AR HDD
Total TLX Score (F(3,69) =39.74, $p < .001$, $\eta^2 = 0.28$)	Control	65.13	41.09	**p < .001*****	**p < .001*****	**p < .001*****
	AR HUD	127.38	44.05	\	p = .46	p = 1
	Cue AR HUD	115.17	45.74		\	p = .26
	AR HDD	129.21	41.14			\
Mental Demand (F(3,69) =20.21, $p < .001$, $\eta^2 = 0.2$)	Control	35.71	24.02	**p < .001*****	**p < .001*****	**p < .001*****
	AR HUD	64.04	23.17	\	p = .5	p = 1
	Cue AR HUD	56.25	23.49		\	p = .5
	AR HDD	63.75	22.92			\
Physical Demand (F(2.42,55.58) =11.33, $p < .001$, $\eta^2 = .14$)	Control	29.42	21.78	**p = .001****	**p = .005****	**p < .001*****
	AR HUD	63.33	24.96	\	p = 1	p = 1
	Cue AR HUD	58.92	25.99		\	p = .71
	AR HDD	65.46	22.07			\
Temporal Demand (F(2.25,51.76) =22.09, $p < .001$ $\eta^2 = 0.28$)	Control	**21.25**	**18.79**	**p < .001*****	**p < .001*****	**p < .001*****
	AR HUD	53.67	25.94	\	p = .52	p = 1
	Cue AR HUD	46.00	24.23		\	p = .65
	AR HDD	54.00	26.83			\
Overall Performance (F(2.36,54.36) =14.4, $p < .001$, $\eta^2 = 0.18$)	Control	**34.83**	**23.70**	**p < .001*****	**p = .003****	**p < .001*****
	AR HUD	64.46	26.56	\	p = .38	p = 1
	Cue AR HUD	58.04	28.51		\	p = 1
	AR HDD	63.29	26.74			\
Effort (F(3,69) =12.84, $p < .001$, $\eta^2 = 0.18$)	Control	**16.54**	**19.87**	**p = .005****	**p = .003****	**p < .001*****
	AR HUD	43.67	34.19	\	p = 1	p = .35
	Cue AR HUD	34.75	29.06		\	**p = .023***
	AR HDD	53.13	32.10			\
Frustration (F(2.18,50.17) =38.82, $p < .001$, $\eta^2 = 0.28$)	Control	**29.42**	**21.78**	**p < .001*****	**p < .001*****	**p < .001*****
	AR HUD	63.33	24.96	\	p = .95	p = 1
	Cue AR HUD	58.92	25.99		\	p = .32
	AR HDD	65.46	22.07			\

Table D.1: A table showing the full list of repeated measures ANOVA and pairwise comparisons for the NASA TLX ratings in Experiment 4.

Experiment 4 Simulator Sickness Model Tables

Timepoint	Mean Total SSQ Score	SD	Min	Max
Baseline	5.691304	11.15576	0	52.36
1	6.389167	9.453765	0	33.66
2	8.570833	13.0716	0	44.88
3	9.973333	12.41235	0	41.14
4	13.24583	20.15498	0	74.8

Table D.2: A table showing a summary of Total SSQ Scores in Experiment 4

Model Intercept	Baseline			T1			T2			T3			T4		
	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig	Est	SE	Sig
Baseline	5.33	2.77		1.06	2.33	= .95	7.92	2.33	p = .17	4.65	2.33	**p = .49***	7.92	2.33	**p < .001*****
T1				6.39	2.75	\	2.18	2.3	p = .35	3.58	2.3	p = .12	6.86	2.3	**p = .004****
T2							8.57	2.75	\	1.4	2.3	p = .54	4.68	2.3	**p = .045***
T3										9.97	2.75	\	3.27	2.3	p = .16
T4													13.27	2.75	

Table D.3: A table showing the Model Estimates, Standard Error (SE) and p values obtained through Wald's approximation for each of the LME models of Total SSQ scores at each timepoint for Experiment 4.

Appendix E

Experiment 5: Analysis Tables

Experiment 5 NASA TLX Ratings Analysis Table

NASA TLX Scale	Condition	Score		Comparison		
		Mean	SD	Visual	Agent	Social
Total ($F(3,69) = 10.5$, $p < .001$, $\eta^2 = 0.15$)	Control	87	45.5326	**$p < .001$*** **	**$p = .021$* **	**$p < .001$*** **
	Visual	127.67	39.38	\	$p = 1$	$p = 1$
	Agent	120.21	36.82		\	$p = 1$
	NDRT only	123.33	33.85			\
Mental Demand ($F(2.34,53.75) = 10.54$, $p < .001$, $\eta^2 = 0.16$)	Control	43.13	25.45	**$p < .001$*** **	**$p = .042$* **	**$p < .001$*** **
	Visual	65.63	20.81	\	$p = .98$	$p = 1$
	Agent	59.75	20.04		\	$p = 1$
	NDRT only	63.58	16.75			\
Physical Demand ($F(3,63) = 3.36$, $p = .024$, $\eta^2 = 0.03$)	Control	43.88	27.47	$p = .052$	$p = .122$	$p = .236$
	Visual	62.04	20.81	\	$p = 1$	$p = 1$
	Agent	60.46	21.85		\	$p = 1$
	NDRT only	59.75	21.19			\
Temporal Demand ($F(3,66) = 7.71$, $p < .001$, $\eta^2 = 0.1$)	Control	**33.09**	**23.15**	**$p = .002$** **	**$p = .043$* **	**$p = .004$** **
	Visual	51.79	23.17	\	$p = 1$	$p = 1$
	Agent	48.46	23.22		\	$p = 1$
	NDRT only	48.13	27.10			\
Overall Performance ($F(3,69) = 6.84$, $p < .001$, $\eta^2 = 0.1$)	Control	**32.21**	**25.06**	**$p = .014$* **	**$p = .131$**	**$p = .004$** **
	Visual	51.38	21.37	\	$p = 1$	$p = 1$
	Agent	45.54	23.97		\	$p = 1$
	NDRT only	50.50	21.35			\
Effort ($F(3,63) = 5.98$, $p = .001$, $\eta^2 = 0.09$)	Control	**27.00**	**26.95**	**$p = .548$**	**$p = .756$**	**$p = .016$* **
	Visual	36.71	22.15	\	$p = 1$	$p = .089$
	Agent	33.92	25.20		\	$p = .085$
	NDRT only	46.09	26.43			\
Frustration ($F(3,69) = 5.98$, $p = .001$, $\eta^2 = 0.1$)	Control	**43.88**	**27.47**	**$p = .028$* **	**$p = .041$* **	**$p = .023$* **
	Visual	62.04	20.81	\	$p = 1$	$p = 1$
	Agent	60.46	21.85		\	$p = 1$
	NDRT only	59.75	21.19			\

Table E.1: A table showing the full list of repeated measures ANOVA and pairwise comparisons for the NASA TLX ratings in Experiment 5.

Experiment 5 RoSAS Ratings Analysis Table

Factor Grouping	Rating	Sig
Warmth	**Warmth (Overall) (F(1,23) = 9.69, p = .005**, η^2 = 0.15)**	**p = .005****
	Happy (F(1,23) = 9.24, p = .005, η^2 = 0.17)**	**p = .005****
	Feeling (F(1,23) =2.74, p = .011, η^2 = 0.05)	p = .011
	Social (F(1,23) = 5.18, p = .032*, η^2 = 0.06)	**p = .032***
	Organic (F(1,23) = 0.54, p = .47, η^2 = 0.006)	p = .47
	Compassionate (F(1,23) = 4.86, p = .038*, η^2 = 0.06)	**p = .038***
	Emotional (F(1,23) = 4.67, p = .041*, η^2 = 0.07)	**p = .041***
Competence	Competence (Overall) (F(1,23) = 1.77, p = .2, η^2 = 0.03)	p = .2
	Capable (F(1,23) =3.04, p = .094, η^2 = 0.05)	p = .94
	Responsive (F(1,23) =4.92, p = .037*, η^2 = 0.08)	**p = .037***
	Interactive (F(1,23) = 1.61, p = .22, η^2 = 0.02)	p = .22
	Reliable (F(1,23) = 2.69, p = .12, η^2 = 0.03)	p = .12
	Competent (F(1,23) = 1.73, p = .2, η^2 = 0.02)	p = .2
	Knowledgeable (F(1,23) = 0.58, p = .45, η^2 = 0.007)	p = .45
Discomfort	Discomfort (Overall) (F(1,23) =0.91, p = .35, η^2 = 001)	p = .35
	Scary (F(1,23) =1.48, p = .24, η^2 = 0.02)	p = .24
	Strange (F(1,23) =1.87, p = .18, η^2 = 0.04)	p = .18
	Awkward (F(1,23) = 0.82, p = .37, η^2 = 0.008)	p = .37
	Dangerous (F(1,23) = 0.18, p = .68, η^2 = 0.002)	p = .68
	Awful (F(1,23) = 3.96, p = .059, η^2 = 0.07)	p = .059
	Aggressive (F(1,23) = 0.06, p = .8, η^2 = 0.001)	p = .8

Table E.2: Full Table showing all the repeated measures ANOVA statistical comparisons of the RoSAS scale for Experiment 5 between the Visual and Agent Conditions.

Appendix F

Experiment 6: Analysis Tables

Experiment 6 NASA TLX Ratings Analysis Table

NASA TLX Scale	Condition	Score		Comparison		
		Mean	SD	Visual	Agent	Social
Total ($F(3,66) = 5.81$, $p = .001$, $\eta^2 = 0.06$)	Control	99.22	39.95	$p = .55$	**$p < .001$*****	$p = 1$
	Visual	111.78	36.42	\	$p = .188$	$p = 1$
	Agent	125.17	35.03		\	$p = .07$
	Social	108.74	37.93			\
Mental Demand ($F(3,66) = 4.99$, $p = .004$, $\eta^2 = 0.07$)	Control	51.91	23.51	**$p = .071$**	**$p = .004$****	$p = 1$
	Visual	60.48	21.52	\	$p = .834$	$p = 1$
	Agent	66.87	17.44		\	**$p = .046$***
	Social	57.13	18.50			\
Physical Demand ($F(2.21,46.42) = 2.51$, $p = .087$, $\eta^2 = 0.02$)	Control	47.30	20.33	$p = .058$	$p = .768$	$p = 1$
	Visual	51.30	20.01	\	$p = .636$	$p = .726$
	Agent	58.30	19.11		\	$p = 1$
	Social	51.61	21.90			\
Temporal Demand ($F(3,63) = 6.35$ $p < .001$, $\eta^2 = 0.1$)	Control	**31.36**	**22.58**	**$p = .033$***	**$p = .009$****	$p = .672$
	Visual	45.39	24.47	\	$p = .978$	1
	Agent	50.83	26.45		\	$p = .107$
	Social	40.61	24.64			\
Overall Performance ($F(3,66) = 9.25$, $p < .001$, $\eta^2 = 0.18$)	Control	**27.91**	**14.90**	**$p = .006$****	**$p < .001$*****	**$p = .005$****
	Visual	43.13	14.82	\	$p = 1$	$p = 1$
	Agent	48.96	18.86		\	$p = 1$
	Social	43.87	18.49			\
Effort ($F(3,66) =5.01$, $p = .003$, $\eta^2 = 0.09$)	Control	**19.52**	**19.16**	**$p = .016$***	**$p = .01$****	**$p = .464$**
	Visual	37.78	26.73	\	$p = 1$	$p = 1$
	Agent	37.96	27.20		\	$p = .828$
	Social	30.70	26.55			\
Frustration ($F(3,66) =3.73$, $p = .015$, $\eta^2 = 0.04$)	Control	**47.30**	**20.33**	$p = 1$	**$p = .003$****	$p = 1$
	Visual	51.30	20.01	\	$p = .131$	$p = 1$
	Agent	58.30	19.11		\	$p = .397$
	Social	51.61	21.90			\

Table F.1: A table showing the full list of repeated measures ANOVA and pairwise comparisons for the NASA TLX ratings in Experiment 6.

Experiment 6 RoSAS Ratings Analysis Table

Factor Grouping	Rating	Comparison	Agent Cue	Social Cue
Warmth	**Warmth (Overall) (F(2,46) = 14.07,p < .001***, $\eta^2 2 = 0.15$)**	**Visual Cue**	**p = .004****	**p < .001*****
		Agent Cue		p = .756
	Happy (F(2,46) = 4.17, p = .022, $\eta^2 2 = 0.04$)	Visual Cue	p = .077	p = .141
		Agent Cue		p = 1
	Feeling (F(2,46) =5.72, p = .007*, $\eta^2 2 = 0.09$)	**Visual Cue**	p = .068	**p = .016***
		Agent Cue		p = 1
	Social (F(2,46) =11.23, p < .001*, $\eta^2 2 = 0.13$)**	**Visual Cue**	**p = .007****	**p = .002****
		Agent Cue		p = 1
	Organic (F(2,46) =1.24, p = .3, $\eta^2 2 = 0.02$)	Visual Cue	p = .771	p = .468
		Agent Cue		p = 1
	Compassionate (F(2,46) = 10.48, p < .001*, $\eta^2 2 = 0.1$)**	**Visual Cue**	**p = .004****	**p = .001****
		Agent Cue		p = 1
	Emotional (F(2,46) =16.21, p < .001*, $\eta^2 2 = 0.24$)**	**Visual Cue**	**p < .001*****	**p < .001*****
		Agent Cue		p = .15
Competence	Competence (overall) (F(2,46) = 0.91, p=0.41, $\eta^2 2 = 0.008$)	Visual Cue	p = 1	p = .735
		Agent Cue		p = .816
	Capable (F(2,46) =1.92, p = .16, $\eta^2 2 = 0.02$)	Visual Cue	p = 1	p = .33
		Agent Cue		p = .354
	Responsive (F(2,46) =2.27, p = .12, $\eta^2 2 = 0.03$)	Visual Cue	p = .306	p = .262
		Agent Cue		p = 1
	Interactive (F(2,46) =0.43, p = .65, $\eta^2 2 = 0.006$)	Visual Cue	p = 1	p = 1
		Agent Cue		p = 1
	Reliable (F(2,46) =4.06, p = .024*, $\eta^2 2 = 0.04$)	**Visual Cue**	**p = .017***	p = .201
		Agent Cue		p = 1
	Competent (F(2,46) =0.09, p = .91, $\eta^2 2 = 0.0009$)	Visual Cue	p = 1	p = 1
		Agent Cue		p = 1
	Knowledgeable (F(2,46) =0.05, p = .95, $\eta^2 2 = 0.0004$)	Visual Cue	p = 1	p = 1
		Agent Cue		p = 1
Discomfort	Discomfort (Overall) (F(2,46)=1.23, p = .3, $\eta^2 2 = 0.02$)	Visual Cue	p = 1	p = .393
		Agent Cue		p = .498
	Scary (F(1.46,33.59) =1, p = .35, $\eta^2 2 = 0.02$)	Visual Cue	p = 1	p = .798
		Agent Cue		p = 1
	Strange (F(1.45,33.28) =1.19, p = .3, $\eta^2 2 = 0.02$)	Visual Cue	p = 1	p = 1
		Agent Cue		p = .184
	Awkward (F(1.5,34.5) =3.17, p = .35, $\eta^2 2 = 0.02$)	Visual Cue	p = .239	p = 1
		Agent Cue		p = .134
	Dangerous (F(1.39,31.93) =0.97, p = .36, $\eta^2 2 = 0.01$)	Visual Cue	p = .852	p = 1
		Agent Cue		p = 1
	Awful (F(1.38,31.84) = 0.06, p = .89, $\eta^2 2 = 0.0008$)	Visual Cue	p = 1	p = 1
		Agent Cue		p = 1
	Aggressive (F(1.48,34.02) =0.72, p = .45, $\eta^2 2 = 0.01$)	Visual Cue	p = 1	p = .984
		Agent Cue		p = .699

Table F.2: Full Table showing all the repeated measures ANOVA statistical comparisons of the RoSAS scale for Experiment 6 between the Visual Cue, Agent Cue and Social Cue conditions.

"What Happens Next?"