Towards a game-based perceptual training tool for improving motorcycle detection

Pancy Poon

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University
for the degree of

Doctor of Philosophy

September 2020

Acknowledgments

Numerous people have helped me along the journey of completing this research and PhD. I am extremely grateful to my supervisor, *Professor David Crundall,* for your support and guidance throughout the PhD. Thank you for believing in my abilities and for affording me the independence to develop my ideas, as well as my Unity skills, but more importantly, thank you for being there to point me in the right directions. Your insightful knowledge and experience have been invaluable. The opportunities that you have provided throughout this PhD has helped me grow tremendously as a researcher.

I am also grateful for my co-supervisors, *Dr Kate Roberts* and *Dr Angela Young*. The constructive advice and guidance that you have given has been instrumental in the completion of my PhD. *Kate*, thank you for giving me objective feedback on my write ups and practical advice about data parsing and study designs. *Angie*, thank you for sharing your knowledge on commentary training with me and for always checking on my well-being.

Thank you to everyone who took part in my studies and who contributed in any way:

- the THINK bike training community, *Tracey* and *Barry*, for accommodating me and helping advertise my studies,
- Devitt Insurance, *Tom*, for saving me a ticket to Motorcycle Live and providing the space for data collection.

I would also like to extend my gratitude to NTUR, in particular *Dr Mark Andrews* and *Dr Jens Roeser* for answering my unrelenting questions regarding R and multilevel modelling, as well as Reece for your help with Unity.

My gratitude also goes out to my examiners *Professor David Shinar* and *Dr Andrew Dunn* for agreeing to review my thesis.

Thank you NTU for funding me, the technicians in Taylor, and a special thanks to *Dr Dawn Watling* who first got me interested in a postgraduate research degree.

Last but not the least, my biggest thanks to God, my family and friends for all the support throughout my studies. Many thanks to my parents *Kieme* and *Katina* for your unwavering love and support, my aunt and uncle *Lewina* and *David* for your love and care. To my partner *Charlie*, thank you for being there for me, I am so blessed to be able to go through this journey with you. Finally, a special shoutout to *Kate* for being such a supportive friend during lockdown!

Abstract

Motorcyclists are disproportionately vulnerable to perceptual failures by other drivers on the road, which are often termed as look-but-fail-to-see collisions. There is theoretical basis to suggest that dual drivers may have an advantage for perceiving motorcycles because they are more familiar with motorcycles. Given the high prevalence of such perceptual failures, there is also scope to explore whether car drivers' perceptual abilities for motorcycles could be enhanced through training. The two main research questions in this thesis are: 1) whether dual drivers have better motorcycle perceptual skills than car drivers and 2) whether car drivers' motorcycle perceptual skills can be improved through a Pelmanism-based motorcycle training tool.

To address the first research question, Experiments 1-3 reported in Chapter 3 compared dual drivers' and car drivers' motorcycle perception on a range of perceptual tasks. No motorcycle-specific perceptual advantage for dual drivers was found. The findings challenge the notion that dual drivers have enhanced perceptual skills for motorcycles than car drivers.

The second research question was addressed in Experiments 4-8. Experiments 4 and 5 demonstrated that dual drivers consistently performed better in terms of matching accuracy on the motorcycle Pelmanism task than car drivers. This is the first piece of evidence indicating that the motorcycle Pelmanism task taps into skills that reflect real world experience. Experiments 6 and 7 then examined the effectiveness of game elements and helped inform the number of rounds to include in the final version of the motorcycle Pelmanism training task that was evaluated in Experiment 8. Unfortunately, this version of the motorcycle Pelmanism training task was not effective in improving drivers' motorcycle perceptual skills, measured by the three perceptual tasks reported in Chapter 3. However, the use of a motorcycle Pelmanism training tool should not yet be ruled out, as research on the motorcycle Pelmanism training tool is still at a preliminary stage. If successful, the motorcycle Pelmanism training tool offers multiple advantages over existing hazard perception training, such as the ability to engage drivers and its high accessibility.

In conclusion, my findings highlight the need to re-evaluate the common assumption that dual drivers are the 'gold standard' in motorcycle perception and contribute to advancing our understanding of a game-based training intervention aimed at improving drivers' motorcycle perceptual skills.

Table of Contents

List of tables

List of figures

Chapter 1 – Introduction to perceptual failures

Chapter overview

The first chapter reviews literature on look-but-fail-to-see errors and the possibility of a dual driver perceptual advantage for motorcycles. The chapter begins by highlighting the vulnerability of motorcyclists according to statistics and the types of collisions that motorcyclists are most vulnerable to. The chapter then focuses on a unique type of collision that disproportionately affects motorcyclists known as look-but-fail-to-see collisions, which are often caused by perceptual failures from the driver colliding with the motorcyclist. The characteristics of look-but-fail-to-see collisions and the underlying factors are described, including bottom-up and top-down factors. The top-down influence of motorcycle familiarity on perception is of particular interest, because there is a group of drivers who are ostensibly superior in their interactions with motorcycles. The potential of motorcycle familiarity as a protective factor against perceptual failures involving motorcycles is thus reviewed.

1.1 Road traffic collisions

1.1.1 Statistics

According to the World Health Organisation (2018), road traffic collisions are the global leading cause of death among people aged between 5-29 years and the eighth leading cause of death globally across all age groups in 2016. In England, road traffic collisions are consistently in the top five leading causes of death among people below 35 years of age (Public Health England, 2018). Although the number of fatalities and severe injuries in Great Britain have been declining, the rate of decline has plateaued since 2010 (Department for Transport, 2019a). With increasing motorisation and population growth, there is a need to develop more effective road safety interventions and training programmes to bring about a reduction in road traffic collisions. Furthermore, injuries from road traffic collisions are costly in terms of treatment, rehabilitation and loss of economic productivity for those who become permanently disabled. The Department for Transport (2019b) estimated that each fatal collision in the UK cost roughly £2.2 million in 2018. Improving road safety for vulnerable road users like motorcyclists is important for reducing road traffic collisions, because they make up 20% of all fatal collisions in Great Britain despite only accounting for 1% of British motor vehicle traffic (Department for Transport, 2019a; Department for Transport, 2019c).

1.1.2 Vulnerable road users – Motorcyclists

Collision statistics from Great Britain show that after taking into account the number of miles travelled by each road user group, motorcyclists emerge at the top with the highest casualty and fatality rate among vulnerable road user groups such as cyclists and pedestrians (Department for Transport, 2019a). In 2018, the rate of motorcycle casualties is approximately 25 times that of cars and the rate of motorcycle fatalities is 60 times that of cars (Department for Transport, 2019a). This pattern of statistics is also reflected in many other countries such as New Zealand and the United States (Ministry of Transport, 2020; National Highway Traffic Safety Administration, 2018). The casualty and fatality rates suggest that motorcyclists are more vulnerable than other road users following a collision. This is likely because they are not protected by a vehicle body and have a higher centre of gravity than other vehicles. Furthermore, motorcyclists are travelling at faster speeds as compared to other vulnerable road users such as cyclists. The average cost of motorcycle casualties resulting from several aspects such as medical costs, material damage and loss of economic output also tends to be higher than other road users (Department for Transport, 2014). An examination of medical costs incurred by motorcycle crashes in Ontario found that injuries sustained from motorcycle collisions were more severe and costly to treat (Pincus et al., 2017). Therefore, motorcyclists are vulnerable road users who are more likely to be involved in collisions resulting in more severe consequences than other road users.

As noted above, motorcyclists are more likely to be killed or seriously injured in a collision than car drivers, but is it solely because collisions involving motorcycles tend to be more severe than collisions involving other vehicles? Looking at collision rates alone, motorcyclists are ten times more likely to be involved in a collision of any severity than cars (Department for Transport, 2019d). The underlying problem of motorcylist vulnerability is therefore that they are overrepresented in traffic collisions. While the severity of motorcycle collisions and the costs incurred may be an incentive to improve motorcycle safety, minimising the severity does not address the underlying problem. In contrast, reducing the number of motorcycle collisions may be more effective for motorcycle safety than minimising the severity of motorcycle collisions. Identifying the type of collisions that motorcyclists are typically involved in can therefore help shape safety interventions that are targeted at reducing motorcycle collision rates.

1.1.3 Types of motorcycle collisions

Using police accident records, Clarke et al. (2004) identified the three most common types of motorcycle collisions. The most prevalent type of collision is right-of-way-violations (ROWV), which occur when a vehicle causes a collision by infringing upon another vehicles' right-of-way (although there is no legal right-of-way in the UK, ROWV is often used in research literature to indicate the failure to yield appropriately). In a typical ROWV involving a motorcycle and a car, the car driver fails to yield appropriately and subsequently turns (or pulls out) into the path of a motorcyclist travelling straight ahead (Transport for London, 2015; Ministry of Transport, 2020). Several examples of how this may occur are shown in Figure 1.1. The second most prevalent type of motorcycle collision results from the loss of control by the motorcyclist, such as losing control when negotiating a bend. This is followed by motorcycle manoeuvrability collisions, which occur because drivers are caught off guard by manoeuvres that only motorcycles are capable of making, such as filtering or taking up unexpected lane positions. For instance, a motorcyclist travelling behind one's car may not be riding in the centre of the lane, where one would otherwise expect a following vehicle to be, but instead be riding to the left or right of the lane.

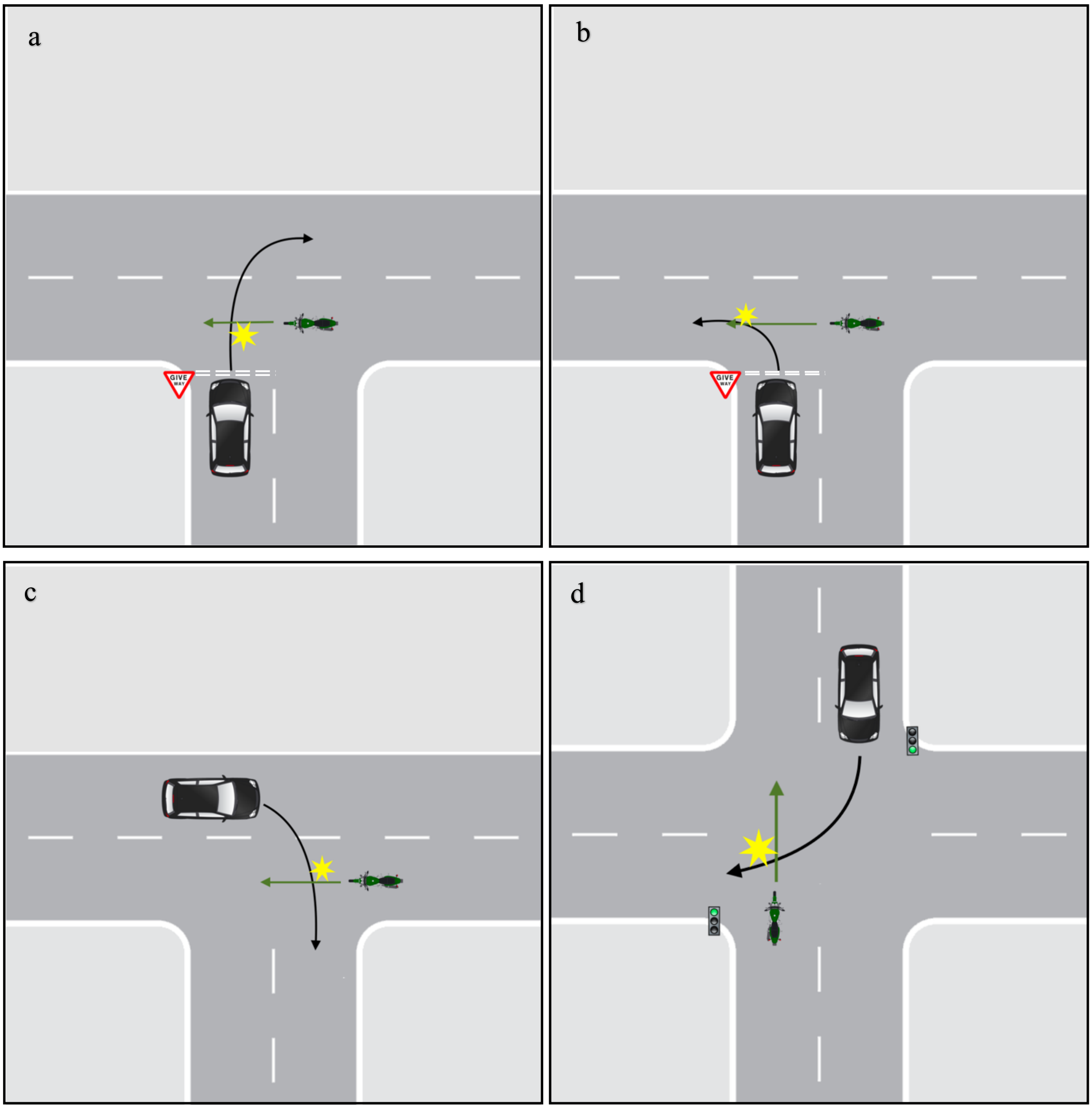

Figure 1.1: Possible ROWV include: a) a right-turning car or b) a left-turning car emerging from a side junction and colliding with a motorcycle travelling straight. c) Car turning into side road or d) Car turning into intersection collides into motorcycle approaching from opposite direction.

Collisions resulting from loss of control usually involve only the motorcyclist but ROWV and motorcycle manoeuvrability collisions often involve another vehicle, the majority of which are cars (Department for Transport, 2015a). In ROWV and motorcycle manoeuvrability collisions, the other driver is more likely to be considered at fault than the motorcyclist

(Clarke et al., 2004). This is supported by an in-depth analysis of motorcycle collisions in Europe, which found that the other driver was responsible for more than 50% of motorcycle collisions whereas the motorcyclist was responsible for only 37% of the cases (ACEM, 2009; see also Hurt et al., 1981). Therefore, while motorcyclists can undertake advanced skills training to help prevent loss of control collisions, there are limited preventative actions for the majority of the collisions that they are involved in. Some of these preventative actions may be weaving laterally when approaching side junctions to increase visibility or riding close to the centre line to stay further away from the junction (e.g., Crundall et al., 2013; Motorcycle Action Group, 2006). However, while motorcyclists may learn to be more cautious around junctions, the onus should not be on motorcyclists to reduce their involvement in collisions caused by other drivers. Similarly, it is unreasonable to expect motorcyclists to stop filtering or perform other manoeuvres unique to motorcycles in order to accommodate other drivers. This is because such manoeuvres are legal and the ability to get through traffic is a common motive for riding motorcycles.

Unfortunately, the lack of consideration from car drivers is a real challenge in enhancing the safety of motorcyclists on the road. A qualitative analysis conducted by Musselwhite et al. (2012) revealed that car drivers viewed motorcyclists as misusing or exploiting the road space because they go against the norms of road use by filtering through traffic. In addition, car drivers were found to place the blame exclusively on riders for being in their blind spots and were quick to interpret behaviours like revving as aggressive. However, motorcyclists may resort to such behaviours to protect their space and signal their presence in response to drivers failing to look and check their blind spots. From a motorcyclist's perspective, an 'aggressive' riding style may be justified given the frequency of collisions due to failure to look properly by the car driver (Department for Transport, 2015a). The lack of consideration for motorcyclists by car drivers is problematic because these drivers are less likely to learn from experiences of collisions or near-collisions with motorcycles and fail to adjust their hazard perception strategies appropriately to avoid future collisions with motorcycles. Instead, they may attribute the collision or near-collision to the motorcyclist and place the responsibility of avoiding future collisions solely on motorcyclists. This was evident in Musselwhite et al.'s

(2012) study, in which car drivers believed that motorcyclists should be held responsible for being involved in collisions. Therefore, the lack of tolerance for motorcyclists may indirectly influence drivers' behaviour when searching for hazards on the road and perpetuate ROWV or motorcycle manoeuvrability collisions.

As motorcyclists are particularly vulnerable to collisions with other vehicles, it is important to examine why these collisions occur. A useful framework to understand why other drivers collide into motorcyclists is proposed by Crundall, Clarke et al. (2008). There are three main processes that are critical to avoiding hazards – looking, processing and appraisal. These processes are hierarchical; drivers need to first look at the appropriate direction of the conflicting motorcycle, then process it sufficiently to recognise that it is a motorcycle, and lastly correctly appraise the risk posed by the conflicting motorcycle. Figure 1.2 below illustrates the chain of driver behaviours that represent these three main processes and how they are influenced by top-down and bottom-up factors. Failure at any of these stages could lead to a collision with a motorcycle.

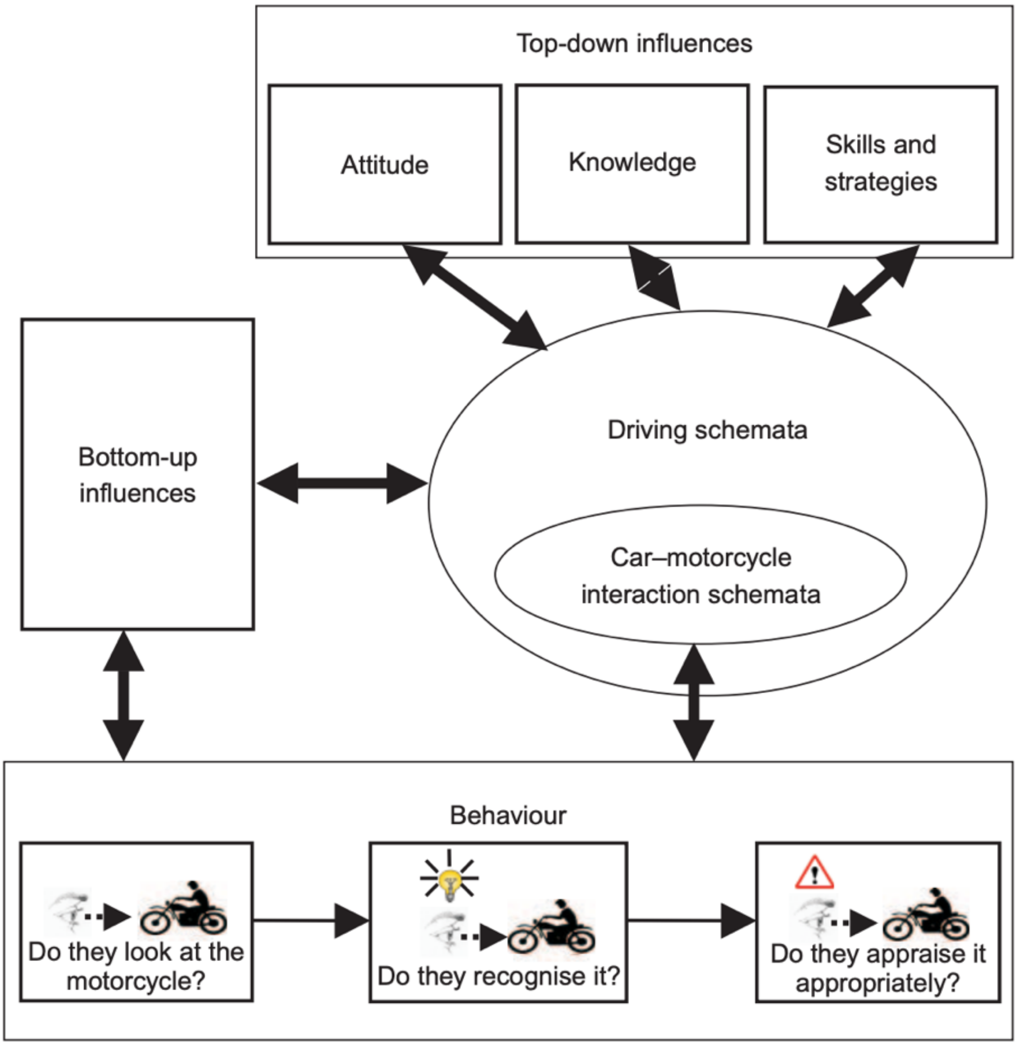

Figure 1.2: Framework describing the three main behaviours for successfully avoiding a motorcycle hazard and the factors that influence these behaviours; adapted from Crundall, Clarke et al. (2008). Permission to reproduce this figure has been granted by the [Open Government Licence](#) for [Crown Copyright](#)

Drivers may fail to look at locations where motorcycles may appear because motorcycles are able to manoeuvre through traffic in ways that are not conventional for cars. An example given in Clarke et al.'s (2004) report described a driver in stationary traffic who failed to check his blind spots or right-wing mirror before making a right turn, thus colliding into a motorcyclist filtering through traffic. Therefore, motorcyclists may be vulnerable to collisions caused by other drivers where those other drivers fail to look in appropriate places before performing a manoeuvre. Where drivers looked at the appropriate location, they may still collide with motorcyclists because they failed to perceive the motorcycle or failed to correctly appraise the level of risk (Crundall, Clarke et al., 2008). The former reason refers to hazard perception whereas the latter refers to hazard appraisal. Although they are two separate processes that each contribute to hazard avoidance, hazard perception tests in the literature often confound hazard perception ability with hazard appraisal. This will be discussed further in Chapter 3. Errors in hazard appraisal could led to risky manoeuvres resulting in a collision. One hazard appraisal error that is pertinent to collisions involving motorcycles is that drivers misjudge the time to collision between their vehicle and the motorcycle. Drivers may incorrectly judge that the motorcyclist is further away or travelling at a slower speed than it actually is (Horswill et al., 2005; Lee & Sheppard, 2017). This is also known as the size-arrival effect, where smaller objects are judged to arrive later than bigger objects because of the slower rate of optical expansion (DeLucia, 1991). Given that motorcycles are smaller than the average car, drivers may be susceptible to misjudge the speed of the motorcycle and the time to collision. As demonstrated in a study, drivers tend to leave smaller gaps when pulling out in front of conflicting motorcycles than conflicting cars (Robbins et al., 2018).

However, the majority of collisions involving motorcycles appear to be due to the drivers' failure to perceive the motorcycle (ACEM, 2009; Clarke et al., 2004, 2007; Department for Transport, 2015a). Perceptual errors occur when drivers fail to detect the presence of a motorcycle or did not process the motorcycle in time to recognise it as a potential hazard. There is also some evidence that drivers are more susceptible to committing perceptual errors than hazard appraisal errors when encountered with motorcycles. Crundall, Humphrey and Clarke (2008) found that under free viewing conditions, drivers were equally accurate in judging the risk of an approaching car compared to a motorcycle. However, when viewing times were limited to 250ms, perception was disrupted and drivers were significantly worse at detecting the presence of a motorcycle than a car. A unique type of perceptual error often reported in collisions involving motorcycles is known as look-but-fail-to-see (LBFTS; Brown, 2002). In these cases, drivers claimed that they had looked but simply had not seen the conflicting motorcycle, despite it being in clear view. The following section reviews the literature surrounding LBFTS collisions, starting with the characteristics of LBFTS collisions and then examining possible factors that contribute to LBFTS collisions.

1.2 Look-but-fail-to-see collisions

1.2.1 What is LBFTS?

Look-but-fail-to-see (LBFTS) errors are commonly defined as the failure to perceive an approaching hazard despite looking in the right direction (Brown, 2002). The most common set up of a LBFTS collision is of a driver emerging from a T-junction and looking to check for conflicting vehicles but failing to see the oncoming motorcycle. There are other situations where LBFTS errors occur, including the ROWVs depicted in Figure 1.1 and during a lane change (although this is also influenced by the presence of blind spots). Currently, much of the literature on LBFTS collisions is limited to collision statistics, which are in turn reliant on drivers' self-report and police assessment of collision causes (see Brown, 2002). As a result, the frequency of perceptual failures may be inflated by drivers' reluctance to admit that they did not look at all or incorrectly judged the level of risk, both of which highlight the driver's negligence (see Crundall et al., 2012). In contrast, reporting a LBFTS error could mitigate the responsibility of the driver in the collision, as it suggests that the collision occurred despite having performed all safety procedures.

However, there is some evidence that LBFTS errors do exist in hazard perception and even in other domains such as reading. For example, the EZ reader model proposes that word identification constitutes an early and late stage of processing (see Reichle et al., 2003). During the early processing stage, the reader attempts to gain a quick impression of the word based on factors such as frequency of the word and the context of the sentence. Upon completion of the early processing stage, the word is further processed while the visual system begins to programme the next saccade. In hazard perception when drivers are scanning the visual field, there is a similar pre-attentive processing stage in which drivers fixate on different areas of the visual field to detect whether a hazard is present. Upon each fixation, visual input is further processed while a saccade is programmed to the next area of the visual field. In both reading and hazard perception, a problem arises if the next saccade is triggered and the eyes move away before the visual input is fully processed. As motorcycles require more effort to process than cars, drivers may be inclined to terminate the processing of a motorcycle prematurely and fail to identify the motorcycle, thus resulting in a LBFTS error. A LBFTS error may also occur if drivers fail to detect the presence of a vehicle entirely, which turns out to be a motorcycle, because perception comprises not only of identification but also detection.

An alternative view of the underlying cause of LBFTS errors is given by Robbins et al. (2019). The authors claimed that LBFTS errors may be due to memory failures rather than perceptual failures. They reasoned that because drivers failed to report the oncoming motorcycle on a high proportion of trials despite fixating on it, this indicated a failure to recall the motorcycle. However, a major flaw in their study was the inability to rule out whether drivers actually processed the motorcycle. There was no evidence that drivers had processed and encoded the motorcycle but subsequently forgot. Given that visual memory tends to be highly accurate for previously attended objects (Hollingworth & Henderson, 2002), it is more likely that drivers failed to process the oncoming motorcycle than failed to remember that a motorcycle was present.

To understand why drivers may look but fail to see motorcycles, an in-depth examination of their perceptual ability for motorcycles is needed. As true LBFTS collisions imply that drivers must have looked in the appropriate direction, some measure of whether they looked at the motorcycle is also needed. Empirical research is thus needed to objectively measure whether drivers looked and processed the motorcycle, in terms of detection and identification, to understand the underlying causes of LBFTS errors in motorcycle collisions. This will be reported in Chapter 3.

1.2.2 Who is involved in LBFTS collisions?

Generally, LBFTS collisions are far more likely to pose a threat to motorcyclists than other motor vehicles (Department for Transport, 2015a). It should be noted that LBFTS collisions are however not exclusive to motorcyclists; previous research has found evidence of LBFTS errors involving bicycles (Department for Transport, 2015b; Summala et al., 1996) and even highly conspicuous police cars (Langham et al., 2002). Nevertheless, the statistics in the previous section suggest that motorcyclists are particularly vulnerable to perceptual failures arising from other drivers. Clarke et al. (2007) reanalysed police accident records and found that the majority of collisions where drivers failed to yield appropriately to a motorcycle could be attributed to their failure to notice the motorcycle, despite it being in clear view. There is also empirical evidence suggesting that drivers are more likely to fail to perceive a conflicting motorcycle than car despite looking in the appropriate direction. Pammer et al., (2018) presented images of a cross junction for 2 seconds and asked participants to report what they saw. In the critical trials, there was an oncoming taxi or motorcycle in addition to an immediate hazard. If participants reported not seeing the additional vehicle, they were then asked to make a forced-choice decision on whether a taxi or motorcycle was present. It was found that participants were twice as likely to be wrong when the additional vehicle was a motorcycle than a taxi, even though there was no difference in participants' estimated likelihood of encountering a taxi or motorcycle. This suggests that drivers may fail to perceive motorcycles when the demands of actual driving are removed, even after controlling for the level of risk and expectation.

In terms of the characteristics of drivers who were responsible for the LBFTS collision, they are often car drivers and tend to be experienced drivers rather than novice drivers (ACEM, 2009; Brown, 2002; Department for Transport, 2015a). This could be due to a range of possible reasons such as over-learned visual search strategies and expectations that disfavour motorcycles (this will be discussed in detail later). Evidence comes from Crundall et al.'s (2012) study, in which they presented a series of video clips displayed across three screens and measured drivers' eye movements. They found that experienced drivers exhibited inappropriate gaze patterns in the presence of conflicting motorcycles and the fixation durations on the motorcycle were short, suggesting insufficient processing. While experienced drivers may be prone to comitting LBFTS errors, those who ride motorcycles may be protected from such perceptual errors because they are found to cause fewer collisions with motorcycles overall (Magazzù et al., 2006). There are several potential reasons for this observation. Drivers who also ride motorcycles (dual drivers) have more experience with and greater exposure to motorcycles than car drivers who do not ride. Consequently, dual drivers may have developed appropriate visual search strategies for motorcycles because they know where motorcycles are likely to appear on the road, or they may have better perceptual skills for spotting motorcycles than car-only drivers. In particular, compared to car-only drivers, dual drivers may have a better search template for motorcycles such that motorcycles stand out easily during visual search, or they may have lower processing thresholds for motorcycles such that they can detect and identify motorcycles quickly. However, no research has been conducted to my knowledge that investigates why dual drivers are less likely to cause a collision with a motorcycle than car-only drivers.

1.2.3 Where and when do LBFTS collisions happen?

A body of evidence suggests that LBFTS collisions most commonly occur at junctions with give-way or stop signs and even uncontrolled junctions where priority-to-the right applies, such as roundabouts or 4-way intersections. Often, a merging or turning vehicle collides into a motorcycle travelling straight on, as described previously. In 2018, statistics collected by the Department for Transport UK (2019e; see also Lynam et al., 2001) reported that 64% of motorcycle collisions occurred at a junction. Pai and Saleh (2008; see also ACEM, 2009) noted that junctions with give-way or stop signs are frequent sites of collisions between motorcycles and automobiles, even more so than uncontrolled junctions. Similarly, Clarke et al. (2007) also found that collisions involving motorcycles were three-times as likely to occur at T-junctions than at uncontrolled junctions.

Collisions at give-way or uncontrolled junctions are not only the most common sites of motorcycle collisions, but they also tend to be more severe than other collision sites. Using a regression model, Pai and Saleh (2008) found that give-way and uncontrolled junctions were a significant predictor of serious injuries sustained by motorcyclists, especially where the motorcycle is travelling straight and is collided into by a vehicle from the opposite direction turning across the path of the motorcyclist (i.e., approach turn collision). Approach turn collisions have the second highest rate and severity of injuries to the head, lower and upper extremities, after head-on collisions with another vehicle (Peek-Asa & Kraus, 1996). Unfortunately, as most motorcycle collisions result from another vehicle failing to yield appropriately, motorcyclists are limited in their ability to safeguard themselves at give-way or uncontrolled junctions. Therefore, research into preventing LBFTS collisions is important to tackle a major problem faced by motorcyclists on the road.

Adverse driving conditions have the potential to significantly restrict the amount of perceptual information available, thus it is useful to consider how they could contribute to LBFTS errors. For example, the risk of a collision between a motorcycle and a car failing to yield increases under low-light conditions such as dusk or dawn hours as well as at junctions found on roads where the speed limit is over 40mph (Pai et al., 2009). However, critics may suggest that the increased risk of collisions may also be due to other factors. For instance, motorcyclists who choose to ride during early or late hours tend to be for sensation-seeking purposes and are likely to be travelling at much higher speeds (see Vlahogianni et al., 2012). While this subset of riders perpetuates the negative stereotype that motorcyclists are reckless and risky road users, they do not undermine the fact that the majority of motorcycle collisions are due to the fault of the other driver. Poor weather conditions can also limit visibility during driving, yet they are not a consistent contributory factor of collisions where a car failed to yield appropriately to a motorcycle (Pai et al., 2009). Clarke et al. (2007) even reported a high occurrence of such collisions in fine weather and dry road conditions. Although adverse driving conditions may heighten the risk of car-motorcycle collisions, they are not as influential as previously thought. The evidence for adverse driving conditions as an underlying cause of LBFTS collisions is therefore not robust enough.

1.2.4 Why do LBFTS collisions occur?

Bottom-up factors – size

There are several reasons as to why motorcycles are more likely to be overlooked than other vehicles. One of the most apparent reasons is that motorcycles are smaller in size than cars – they have a face-on silhouette that is 30-40% that of a passenger car's, thus making them difficult to spot (see Huang & Preston, 2004). The small size of motorcycles means that they are easily obscured by traffic, which is exacerbated by the fact that if motorcyclists do filter through traffic, they are likely to be obstructed by other vehicles as they are between vehicles. Certainly, when motorcyclists filter through traffic, a mismatch between car drivers' expectations of where a hazard should appear may also contribute to a collision, and this will also be discussed later. As mentioned in the previous section, drivers may be susceptible to the size-arrival effect because of the small size of motorcycle. However, as explained, the size-arrival effect is a hazard appraisal error and describes a scenario where the driver has already successfully perceived the motorcycle (see also Pai, 2011).

Motorcycles are not only smaller in size but also narrower than other vehicles on the road, and the high spatial frequency poses a difficulty for other road users to detect them. The literature on the hierarchical organisation of visual perception has documented a global precedence effect whereby global information is processed before local information (Navon, 1977). This effect is in turn driven by the tendency for low spatial frequency targets to be processed earlier than high spatial frequency targets (Han et al., 2003; Lamb & Yund, 1993). This suggests that motorcycles are not perceived as readily as cars due to the natural tendency of the visual system to favour wider, lower spatial frequency objects like cars over narrower, higher spatial frequency objects like motorcycles.

A phenomenon resulting from the tendency to process wide over narrow objects is motion camouflage, whereby a small object in front of a larger object remains below the detection threshold even though they are in motion (Motorcycle Action Group UK, 2006). Usually, motion can exogenously draw drivers' attention to the motorcycle. However, in this instance, the position of the motorcycle relative to the larger object is constant, which masks the approaching motorcycle because the observer only detects the edge of the larger object and not the smaller object. Empirical evidence from Rogé et al.'s (2010) study showed that drivers struggled to detect motorcycles from a distance particularly when the surrounding traffic consisted of heavy goods vehicles, which are wider and therefore prioritised over the motorcycle. Drivers were able to spot the motorcycle from a greater distance when the surrounding traffic consisted of light vehicles than heavy goods vehicles and this benefit was further enhanced for high-contrast motorcycles (Rogé et al., 2010). The contrast between the background and the motorcyclist is another factor that influences the level of conspicuity. Conspicuity is the ability of an object to attract visual attention as a result of its features.

Bottom-up factors – conspicuity

For a long time, researchers have attributed low motorcycle conspicuity to LBFTS collisions, specifically because their often-dark colours and irregular contours require more effort to detect and process than other vehicles (Underwood, Humphrey, & van Loon, 2011; Williams & Hoffman, 1979; Wulf et al., 1989). Conspicuity of an object is determined by its physical properties such as contrast, luminance and colour. Accordingly, Rogé et al. (2012) showed that enhancing the colour contrast between the motorcyclist and the background increases the distance at which drivers can detect oncoming motorcycles. Research using case-control designs have also shown that increasing the luminosity of the rider by wearing fluorescent clothing or light-coloured helmets is associated with a lower collision risk involving another vehicle (Comelli et al., 2008; Wells et al., 2004). Crucially, colours that contrast with the environment, rather than bright or fluorescent colours per se, act to increase the detectability of motorcyclists (Gershon et al., 2012; Hole et al., 1996). For instance, the usual advantage of detecting brightly coloured motorcyclists was not found in rural settings, as darker clothing created a greater contrast between the more colourful rural environment, resulting in quicker detection (Gershon et al., 2012).

In addition to the contrast between the rider and background, the contrast between the motorcycle body and background can also account for the low conspicuity of motorcycles on the road. Research investigating the use of daytime running lights (DRL) to maximise the contrast between motorcycles and the environment found that DRLs can help reduce motorcycle collisions, especially fatal collisions (Krajicek & Schears, 2010; Radin et al., 1996; Yuan, 2000). However, as other road users are increasingly using headlights during the day, Cavallo and Pinto (2012) noted that the effectiveness of DRLs in improving motorcycle conspicuity has declined. Motorcycle detection rates decreased in the presence of headlights from other road users. In response to DRLs no longer being a unique visual feature of motorcycles, the authors proposed the use of a unique colour or configuration of DRL for motorcycles as an alternative. In a subsequent study, Pinto et al. (2014) compared the effect of different DRL configurations and colour against the standard central white headlight. Adding a white helmet light was found to improve detection of motorcycles that were far away and overall, motorcycles with a central yellow headlight were more likely to be detected than motorcycles with a central white headlight. The authors suggested that changing the colour of the headlight increased the distinction between motorcycle DRLs and other white headlight sources from surrounding traffic, thereby providing a unique way of identifying motorcycles. Similarly, another study in Australia noted the advantages of yellow DRLs for motorcycles in contrast to other vehicles, which are currently not permitted to be fitted with yellow DRLs (Paine et al., 2005). Interestingly, the study by Pinto et al. (2014) did not find any significant improvements in motorcycle detection rates when using a triangular headlight configuration that is different to cars. Although this seems to contradict previous findings that triangular configurations enhance motorcycle conspicuity (e.g., Gould et al., 2012a, 2012b; Cavallo et al., 2015), other studies were actually aimed at improving drivers' judgement of the motorcycle's approaching speed. Therefore, based on the findings discussed here, triangular headlight configurations may be a good countermeasure for the size-arrival effect but its effectiveness in reducing perceptual errors such as LBFTS may be less effective than coloured headlights.

While the low conspicuity of motorcycles is one factor that contributes to drivers' tendency to overlook motorcycles, LBFTS collisions still occur under high visibility conditions and when the motorcyclist was wearing fluorescent clothing or headlights (ACEM, 2009; Clarke et al., 2007). Therefore, modifications to increase the conspicuity of motorcycles may not be a universal solution for perceptual failures leading to LBFTS collisions. Furthermore, empirical evidence suggests that drivers may struggle to detect highly conspicuous police cars despite having reflective stripes, large blue and yellow patterns and flashing lights (Langham et al., 2002). In their study, two identical police cars were presented across different trials, one parked in line with the lane and the other parked at an angle, both of which were equally conspicuous. Experienced drivers detected the police car parked in line consistently slower than the one parked at an angle. As the direction of the police car parked in line is congruent to the direction of traffic, it is likely that that police car did not receive attentional priority for processing. This suggests that conspicuity is not the sole factor in LBFTS collisions and perceptual failures may still occur if attention is not directed to a highly conspicuous object. As attention helps guide perception in a top-down manner, the perception of natural scenes often involves an interaction of bottom-up and top-down processes. Therefore, it is relevant to also understand the top-down factors that contribute to LBFTS errors.

Top-down factors – inattentional blindness

The top-down influence of attention on perception has been widely researched. A prominent theory proposed to explain LBFTS errors is inattentional blindness. Inattentional blindness refers to the failure to notice an unexpected object in the visual field because attention was directed at something else and not towards the overlooked object (Mack & Rock, 1998). Most and Astur (2007) suggested that motorcycles are easily overlooked because they are not as well represented in drivers' attentional set, which determines the items or features that should receive attentional priority. In their study, participants were instructed to follow the directions of either yellow or blue arrows, and by doing so shaped their attentional set for yellow or blue objects. It was found that participants were slower to detect and crashed more often into motorcycles that were incongruent to the colour of the arrows they had been following as compared to congruent motorcycles. Furthermore, Pammer et al. (2018) showed that motorcycles received significantly lower attentional priority than cars, which could explain why LBFTS collisions often involve motorcycles. When an immediate hazard was present, drivers were less likely to process motorcycles than cars.

Change blindness is a similar but distinct perceptual phenomenon from inattentional blindness and is sometimes also used to explain LBFTS errors. Change blindness is the failure to notice a change in two images of a near-identical scene when a brief disruption occurs between the images (Rensink et al., 1997; Simons & Rensink, 2005). It is relevant to driving because brief disruptions or changes to a driver's visual field often occur in the form of head turns or eye movements such as saccades or blinks, and drivers have to maintain a coherent representation of the situation. Imagine a driver who is pulling out from a side road and performing various checks before making a manoeuvre, first checking for oncoming traffic on the right, then checking for a clear turning path before performing a last check for oncoming traffic on the right. If a motorcycle appears when the driver's visual field has shifted during the series of checks, the driver might not have noticed the additional motorcycle approaching from the right, resulting in change blindness. If the motorcycle had been present since the first check and the driver had not yet detected it, change blindness can still occur because the change in optical size of the motorcycle may still fall below the threshold for change detection (see Horswill et al., 2005). However, Galpin et al. (2009) pointed out that laboratory conditions inducing change blindness are more restrictive than real life conditions, thus inflating the occurrence of change blindness. For example, drivers can rely on other cues such as motion to detect changes given that objects, particularly hazards, will continue to move across the visual field. In turn, there may be other top-down factors that can reliably explain LBFTS collisions.

Top-down factors – expectation

Another reason for LBFTS errors may be mismatches between the driver's expectation and the situation. Drivers' expectations are often derived from schemas developed over time and are based on previous experiences of potential hazards in traffic. Schemas are a cognitive framework that help organise and interpret information more efficiently, so naturally schemas that have previously been successful in predicting potential hazards will continue to be used. In other words, drivers' schemas of potential hazards in traffic are shaped by their accumulative experience. In turn, schemas influence the way drivers extract information from the visual scene and this means that rare events or objects are more likely to be neglected than frequently occurring objects (Clarke et al., 2007; Summala et al., 1996). Therefore, the low proportion of motorcycles in traffic lowers drivers' expectations of seeing a motorcycle, and this is also known as the target prevalence effect (Beanland et al., 2015). This effect whereby drivers adjust their expectancy and search strategy according to the frequency of the target has been well documented (Wolfe et al., 2005; Wolfe et al., 2007). Given that motorcycles appear infrequently, a schema that is biased against motorcycles allows drivers to quickly and correctly conclude that a motorcycle is not present in the majority of the cases, and thus is often reused when searching for potential hazards. However, the lowered expectation has an adverse effect on drivers' sensitivity to motorcycles when one does appear because they are less likely to detect or recognise that a motorcycle is present despite having fixated on it. Furthermore, the low motorcycle prevalence effect is exacerbated by the natural tendency to process objects of low spatial frequency over high spatial frequency (Han et al., 2003). Beanland et al. (2015) manipulated the prevalence of buses and motorcycles on a simulated drive and found that drivers detected high prevalent vehicles from further away than low prevalent vehicles, but the effect of prevalence was greater for buses than for motorcycles. This indicates that the physical properties of motorcycles interact with top-down factors such as expectations.

When drivers are scanning the visual field for hazards, an attentional set is activated to guide hazard search based on their schema about potential hazards (see Reeder & Peelen, 2013). If drivers do not expect to encounter a motorcycle, their attentional set is unlikely to include motorcycles, which biases perception against motorcycles. The attentional set therefore reflects the top-down influence of factors, such as expectation following from target prevalence, on visual search. According to the multiple-decision model, two types of decisions are made during visual search – an overall decision of whether the visual scene contains a target and whether the object currently fixated on is a target (Schwark et al., 2013; Wolfe & Van Wert, 2010). The former decision refers to a quitting threshold for hazard search and the latter refers to a threshold for object identification upon fixation. Therefore, the low prevalence of targets can shift the quitting threshold such that visual search terminates prematurely because the target is unlikely to be present. Evidently, Wolfe and Van

Wert (2010) demonstrated that termination of search on target-absent trials becomes quicker as the frequency of target decreases. The identification threshold can also be shifted by low target prevalence such that more visual information is needed to correctly identify the object. Eye-tracking evidence suggests that low prevalence targets are less likely to be correctly identified upon first fixation and take longer to be verified as a target than high prevalence targets (Godwin et al., 2015; Peltier & Becker, 2016).

Given that expectations are derived from schemas that are based on previous experience, experienced car drivers are likely to have a low expectation of motorcycles because they make up less than 1% of Great Britain's motor vehicle traffic (Department for Transport, 2019c). If expectations shaped by real world experience play a role in LBFTS errors, it is expected that experienced car drivers would be more prone to committing LBFTS errors than novice drivers. Previous studies have reported that experienced car drivers are more likely to overlook and collide with motorcycles than novice drivers (Clarke et al., 2007; Van Elslande & Faucher-Alberton, 1997). Using eye-tracking measures, Labbett and Langham (2006) reported that fixations of experienced car drivers tended to be limited to certain areas of the road where hazards normally occur, such as the centre and focus of expansion of the road. In contrast, novice drivers were found to fixate on other parts of the road that are not prioritised by experienced car drivers. Consequently, when a motorcycle appears outside of areas that experienced drivers have learned to expect hazards, such as the centre of the road, novice drivers are able to detect the motorcycle more quickly than experienced car drivers. As Labbett and Langham (2006) used a small sample size and stimuli, Crundall et al. (2012) improved on their methodology by comparing dual drivers', experienced and novice car drivers' eye movements when presented with conflicting vehicles at a T-junction. An analysis of gaze locations during hazard search showed that only the experienced car drivers failed to adjust their search to accommodate an approaching motorcycle. The difference in gaze locations between novice and experienced car drivers indicates that accumulated driving experience biases hazard search against motorcycles, presumably because the low prevalence of motorcycles on the road skews drivers' expectations. Interestingly, dual drivers (drivers who also ride motorcycles) were not only safer in the presence of motorcycles, but their eye movements also demonstrated increased sensitivity to motorcycles as compared to experienced car drivers (Crundall et al., 2012). This suggests that dual drivers may be less susceptible to overlooking motorcycles than experienced car drivers. The difference between the two groups of drivers is that dual drivers have more experience with and exposure to motorcycles than car drivers. Therefore, it is possible that familiarity with motorcycles, derived from the increased experience and exposure, acts as a protective factor against LBFTS errors.

Protective factor – familiarity

In this thesis, familiarity refers to having prior knowledge about the object category after repeated exposure, which is the dictionary definition of familiarity (see also Malt & Smith, 1982). There is some evidence suggesting that familiarity with motorcycles serves as a protective factor against collisions with motorcycles. Magazzù et al. (2006) reported that dual drivers are less likely to collide with motorcycles than car drivers, and this protective effect of familiarity also extends to friends and relatives of dual drivers who have ridden pillion (Brooks & Guppy, 1990, as cited in Crundall et al., 2012). A coherent explanation for both findings is that increased exposure to motorcycles, whether through personal or vicarious experiences, increases familiarity with motorcycles. However, as both of these studies examine the likelihood of colliding with motorcycles, the findings offer limited insight into how familiarity lowers the likelihood of colliding with motorcycles. According to Crundall, Clarke et al.'s (2008) framework for understanding motorcycle collisions, the familiarity advantage could arise at any of the looking, processing or appraisal stages. Previously, Crundall et al. (2012) reported that dual drivers made safer responses than car drivers around motorcycles with regards to the decision to pull out in front of a conflicting motorcycle. This would suggest that increased familiarity with motorcycles benefits drivers' ability to appraise the situation. However, as perceptual failures are most commonly reported and observed in motorcycle collisions (ACEM, 2009; Clarke et al., 2004, 2007; Department for Transport, 2015a), it would be interesting to find out whether familiarity has a protective effect against perceptual failures.

Theoretically, drivers who are familiar with motorcycles are likely to have more accessible representations of motorcycles in the long-term memory than drivers who are unfamiliar with motorcycles. Research has shown that long-term memory representations have a top-down influence on attention and perception (Christie & Klein, 1995; Olivers, 2011). These representations may include prior knowledge of the typical features that are representative of motorcycles, which can be used to bias exogenous attention during hazard search when activated by an attentional set (see Reeder & Peelen, 2013). According to Reeder and Peelen (2013), the attentional set, also known as the search template, facilitates visual search by directing attention towards objects that match the features within the attentional set. As a result, drivers who are familiar with motorcycles may benefit from an attentional set that includes motorcycle features when they are scanning for hazards. Familiarity with motorcycles may also lower the processing threshold for motorcycles (see Curby & Gauthier, 2009). Upon fixation on an object, the visual input is processed to determine what the object is by accumulating evidence or visual information until the threshold for object recognition is reached. For drivers who are familiar with motorcycles, the processing threshold may be lowered such that less visual information is needed to recognise the motorcycle. Based on this reasoning, dual drivers may be less susceptible to terminate visual processing of a motorcycle prematurely by saccading away, as described in a previous section on the EZ reader model.

Preliminary evidence suggests that there may be a familiarity-based advantage for motorcycle perception in dual drivers. As mentioned earlier, eye movement data from Crundall et al.'s (2012) study suggested that dual drivers were more sensitive to the presence of motorcycles than car drivers. They also reported that dual drivers fixated longer on motorcycles than cars whereas this was reversed for car drivers. The authors suggested that dual drivers were aware of the high processing demands of motorcycles at the pre-attentive processing stage and thus directed their attention to processing motorcycles rather than saccading away. In contrast, car drivers fixated on the motorcycle but may have saccaded away before having processed the motorcycle sufficiently to identify it. Therefore, it is possible that dual drivers' familiarity with motorcycles may have enhanced their sensitivity to motorcycle features during visual search and lowered their processing threshold for motorcycles such that they can detect and process the motorcycle in time before saccading away. Further evidence for the protective effect of familiarity against motorcycle perceptual failures comes from a study by Lee et al. (2015). They compared car drivers from the UK, who are expected to be unfamiliar with motorcycles, and Malaysia, who are expected to be familiar with motorcycles due to the high proportion of motorcycles on the road. While British drivers were significantly poorer at detecting motorcycles than cars, this effect was reduced in Malaysian drivers. This suggests that drivers who are familiar with motorcycles may experience a perceptual advantage for motorcycles.

Based on the literature reviewed, it is possible that a familiarity-based perceptual advantage exists. This hints at a motorcycle-specific perceptual advantage for dual drivers because they are more familiar with motorcycles than the average car driver. To examine whether dual drivers have a motorcycle-specific perceptual advantage, a useful approach is to assess perceptual skills based on the looking and processing stages of Crundall, Clarke et al.'s (2008) framework for understanding motorcycle collisions. This thesis chooses to focus on looking and processing because these processes are pertinent to perceptual failures, which are characteristic of LBFTS collisions involving motorcycles. By comparing dual drivers' ability to perceive motorcycles with that of car-only drivers, research can determine whether dual drivers have developed perceptual skills that protect them against LBFTS errors. If a dual driver advantage is found in perceiving motorcycles, training interventions can then look to dual drivers as the gold standard in overcoming perceptual failures involving motorcycles.

1.3. Thesis overview

The first chapter has provided the background research relevant to this thesis, particularly to the first empirical chapter. The introduction described and reviewed existing literature on LBFTS collisions involving motorcycles and discussed the underlying factors that contribute to LBFTS errors in a bottom-up or top-down manner. The possibility of a familiarity-based protective effect for dual drivers against perceptual failures was also discussed based on preliminary evidence. On this account, the first empirical chapter (Chapter 3) will investigate my first research question – whether dual drivers have better motorcycle perceptual skills than car drivers. The first experiment investigates whether dual drivers have better hazard perception for motorcycles than car drivers, using a hazard prediction task that relies on drivers looking at the appropriate locations to spot the developing hazard. The second experiment investigates whether dual drivers are quicker to spot a motorcycle than car drivers using a visual search task and the third experiment investigates the processing thresholds for successful motorcycle detection and identification using a T-junction task. The first empirical chapter (Chapter 3) therefore provides clarification on the general notion that dual drivers are better at perceiving motorcycles than car drivers.

Given the prevalence of LBFTS collisions involving motorcycles, the thesis recognises that the present research needs to go beyond investigating the dual driver advantage. There is a need to also examine preventative methods that improve car drivers' ability to perceive motorcycles. The second research aim of the thesis is thus to further develop an improved version of a game-based motorcycle perceptual training task, which was introduced recently, and to evaluate the effectiveness of this version of the training. The first empirical chapter also sets the foundation for the second part of the thesis because the same perceptual tasks are used to evaluate the training. If the perceptual tasks demonstrated a dual driver advantage and subsequently found a training effect, we can infer that the training increased drivers' familiarity with motorcycles and improved their motorcycle perceptual skills similar to that of the 'gold standard' dual drivers. As the focus of the research shifts to developing a perceptual training tool for car drivers, a second literature review is presented in Chapter 4 after the first empirical chapter. This literature review discussed and evaluated existing hazard perception training interventions and preventative efforts related to motorcycle safety. Key limitations of existing hazard perception training interventions were identified, including the lack of a training that specifically targets motorcycle hazards and the challenge of motivating drivers to participate in the training voluntarily. However, a motorcycle Pelmanism task showed potential to be an effective perceptual training method for car drivers. Therefore, Chapter 4 also reviewed literature surrounding serious games, which are games designed for learning rather than entertainment purposes. This included the role of game elements for enhancing players' motivation and the theoretical bases for a motorcycle perceptual training task inspired by a game of Pelmanism.

Following the literature review on game-based training for motorcycle hazards, the subsequent empirical chapters (Chapters 5 & 6) report a series of experiments to examine the sensitivity of a motorcycle-specific Pelmanism training task and further develop the motorcycle Pelmanism task as a training tool. Chapter 5 investigated whether the motorcycle Pelmanism task reliably captures differences in performance between drivers who are naturally familiar with motorcycles and drivers who are not familiar, without any prior training. Chapter 6 then examined the optimal number of rounds to include in the motorcycle Pelmanism training task and investigated the effects of game elements on players' motivation and performance on the training. These findings ultimately inform the design of the Pelmanism-based motorcycle perceptual training task in the final experiment (Chapter 7), in which its effectiveness was evaluated using perceptual tasks that were reported in Chapter 3. Finally, a general discussion of the findings and implications of the thesis is discussed in Chapter 8.

To summarise, my two main research questions are:

1) Do drivers with motorcycle riding experience (dual drivers) have better motorcycle perceptual skills than car drivers?
2) Can car drivers' motorcycle perceptual skills be improved through a motorcycle Pelmanism perceptual training task?

In the next chapter, the general methods and software used to create stimuli are described and a clear rationale for the type of statistical analysis conducted throughout the experiments in this thesis is provided, before presenting the first empirical chapter.

Chapter 2 – General Methods

This chapter describes the general methods and software used to create the stimuli in this thesis. As the exact methodologies used to create stimuli for each experiment were different, they will be described in further detail when reporting each experiment. The general method of analysis is also explained in this chapter. Apart from Experiment 1, data from each experiment were analysed using linear mixed effect models. As this method of analysis is not yet commonly used within road safety research, this chapter will also rationalise why linear mixed effect models are preferable to traditional analyses such as ANOVA.

2.1 Software used to create stimuli

2.1.1 Adobe Photoshop

The static stimuli used in Experiments 2-8 were created on Adobe Photoshop using images of vehicles or pedestrians, which were either sourced from the internet or taken from photographs captured around Nottinghamshire. In several experiments that required stimuli consisting of a vehicle or pedestrian on an empty road, the general approach was to crop out the vehicle or pedestrian from the original background using the Lasso tool and then to layer it on top of an empty road background. To enhance the realism of the stimuli, a drop shadow was created for the object via the 'layers' tab. Using tools such as Rotate, Skew, and Perspective, the drop shadow was then transformed perpendicular to the object so that it appeared to be a shadow cast onto the ground. In Experiments 2 and 8, the visual search arrays for the visual search task were also created on Adobe Photoshop by cropping traffic scenes to the same sizes and arranging them into a 4x4 array with the help of grid lines available in the software.

2.1.2 Adobe Premiere Pro

The dynamic stimuli used in Experiments 1 and 8 were edited in Adobe Premiere Pro and exported in MPEG-4 (mp4) before being uploaded onto an Oculus Go headset. The export settings were as follows: VBR Pass 1, target bitrate 25, maximum bitrate 75, monoscopic and rendered at maximum bit depth.

2.1.3 E-prime 3.0

The T-junction task in Experiments 3 and 8 was programmed using the E-prime 3.0 software (Psychology Software Tools, 2016). The software includes applications for building experiments (E-studio), running experiments (E-run), as well as for data analysis (E-merge and E-DataAid). To programme the T-junction task, stimuli were imported into E-studio and the correct response for each stimulus was coded so that the experiment automatically recorded whether participants' responses were correct or incorrect. The flow of the experiment was as follows: instructions were presented to participants on the screen, followed by a practice block consisting of 6 trials and an experimental block consisting of 240 trials in Experiment 3 or 50 trials in Experiment 8. The experiment was run on a computer connected to a 17.2" CRT (Cathode-ray tube) monitor with resolution of 1024 X 768 using the E-run application.

2.1.4 Unity

The majority of the experiments were coded on Unity, namely the visual search task and Pelmanism tasks in Experiments 2 and 4-8. Unity is a development platform commonly used to design and develop games. The visual search task was coded in Unity because the use of scripts for experimental building was more flexible than the drag-and-drop method in E-prime. In Unity, it was possible to build and design an experiment according to any specifications whereas the ability to customise the experiment was more limited in E-prime. Unity was also the preferred choice for coding the Pelmanism tasks in Experiments 4-8 in order set up the motorcycle Pelmanism training tool as a serious game. Unity offers a more advanced and contemporary visual experience compared to that in E-prime, allowing the end design of the Pelmanism task to be more convincing and game-like than in E-prime. The visual experience is important for players to feel immersed in the Pelmanism task, similar to when playing entertainment games, and this is relevant to researching Pelmanism as a game-based training tool.

In addition, the experiments built using Unity could be exported either as a web version or a standalone version for Windows or MacOS computers. For standalone versions, participants' game data were simply written as a new file saved to the computer, similar to standard psychology experimental software such as E-prime. For the web version, Unity exports the experiment as a WebGL file and compiles all the code in JavaScript, which allows the experiment to be run online on most browsers. The ability to build an experiment capable of running online was beneficial for my data collection as it allowed me to reach less accessible sample groups, such as dual drivers via online testing. Experiments 4 and 5 were conducted online while the rest of the experiments were conducted face-to-face. For Experiments 4 and 5, the WebGL file was hosted on a free game-hosting platform called simmer.io (https://simmer.io/). Participants' game data were stored on a server called dreamlo (http://dreamlo.com/) and are only accessible through a private URL with a keycode. Online testing was conducted successfully and their shareable links are http://bit.ly/PelmanismFullVer and https://bit.ly/Pelmanism2 respectively.

2.2 General analysis

2.2.1 Why linear mixed effect models?

Linear mixed effect models, also known as multilevel models, have advantages over traditional analyses using general linear models (GLM) such as ANOVA or ordinary least squares regression. In experimental studies, multiple data points are usually measured from an individual participant, for example by presenting multiple trials to the participant. Responses from the same individual are expected to be correlated more than responses between individuals because of individual characteristics such as response criterion or response bias. This means that each data point is nested within participants and are therefore not independent. When data violates the assumption of independent errors (i.e., residuals) required by GLMs, it leads to inflation of Type I and Type II errors (Field & Wright, 2011).

That is, the error of finding a significant effect when there is not one and not finding a significant effect when there is one. To overcome this violation of independence, an average score is usually computed in GLMs, however a lot of data is lost by doing so because much of the within-individual variances would be unaccounted for, which is particularly problematic for repeated observations (see Davidson & Martin, 2013). Signal detection theory is another common approach used to analyse detection rates by calculating the proportion of yes and no responses when a target is present or absent (see Green & Swets, 1966; Macmillan & Creelman, 2005). However, similar to GLMs, responses are aggregated across trials to produce hit rates, miss rates, or false alarm rates for each participant, thereby again disregarding variance within individual participants and variance across stimuli (Rabe, 2018). In contrast, linear mixed effect models take into account all data points while also modelling characteristics of naturally occurring clusters.

Linear mixed effect models are an extension of GLMs, as they also model a linear relationship between predictor variables (or independent variables in ANOVA) and the outcome measure. However, unlike GLMs where the effect of predictor variables is assumed to be the same for different participants or stimuli, linear mixed effect models allow the effect of predictor variables to vary according to grouping variables, which could be by participant and/or stimuli (Hayes, 2006). This is achieved by fitting a unique intercept and/or slope for each grouping variable and they are referred to as random effects. Therefore, while GLMs estimate only a fixed parameter for the effect of predictors, linear mixed effect models include both a random and fixed parameter.

Linear mixed effect models could also model more than one source of random effect simultaneously. In repeated measures designs, the stimuli presented to participants do not differ between conditions. In the same way that responses from the same participant tend to be correlated, the same stimulus is expected to elicit more similar responses across all participants. According to multilevel modelling, this is known as a crossed random effect and can be appropriately modelled using the lme4 package in R (Baayen et al., 2008; R Core Team, 2020). Given that repeated measures designs are often interested in the change in responses within participants, it is important to model the relevant random effects present in the data (Quené & Van den Bergh, 2008). This is because traditional analyses that average responses across stimuli ignore the systematic variation between experimental stimuli, which in turn can lead to massive inflation of Type I errors (Judd et al., 2012). For the findings of this thesis to be generalised beyond the sample of stimuli developed and used in this thesis, it is important to include stimuli as a random effect.

2.2.2 Performing multilevel modelling analysis

Linear mixed effect models are constructed using the lme4 package (Bates et al., 2015) written in the statistical programming language R (R Core Team, 2020). During multilevel modelling, a series of models is constructed and compared. First, a baseline model of the random effects and covariates is constructed. The main effects of interest are then added to the baseline model, followed by the interaction terms between the main effects. Each fixed effect (i.e., main effect or interaction term) can be tested using a likelihood ratio test.

To test the effect of each predictor variables or fixed effects, likelihood ratio tests are used to compare a model with and without the effect. The likelihood ratio test compares the fit of one model (i.e., -2 Log Likelihood or -2LL) with the other and tests whether the difference is statistically significant. The resulting likelihood ratio test statistic is approximately chi-squared with degrees of freedom equal to the number of additional parameters in model A compared to model B (Hayes, 2006). Note that the likelihood ratio tests can also be used to compare general linear models. When there are more than several main effects and interactions of interest, the drop1 function with likelihood ratio test option from the lme4 package can be used to speed up the process. With the drop1 function, separate models for each fixed effect need not be constructed because this step is incorporated within the function. The drop1 function drops each fixed effect in turn from the model containing it and computes the change in model fit, which returns the same likelihood ratio test statistic described above. The drop1 function follows the principle of marginality, meaning that only the highest order terms in the model will be dropped. For example, drop1 applied to a model containing a three-way interaction would only test the three-way interaction because the two-way interactions and main effects are marginal to the three-way interaction. Therefore, where the drop1 function is used during the analysis, the drop1 function is applied to the main effects model and interaction model(s) up to the highest order interaction. The main effects or interaction terms would then be dropped in turn from their respective models.

To compare the levels within each categorical fixed effect, the default coding scheme used is dummy coding, unless otherwise stated in the methods section. Dummy coding compares each level within the categorical variable to a fixed reference level. For example, if we choose group 1 to be the reference level, then groups 2 and 3 will each to compared to group 1. This differs from the effect coding scheme resembling ANOVA, which compares each level to the grand mean of all levels (e.g., mean of groups 1-3).

2.3 Ethics

All studies were carried out in accordance with the Economic and Social Research Council's guidelines for ethical research and approved by the College Research Ethics Committee of Nottingham Trent University. Informed consent was obtained from all participants prior to their participation. All data collected from participants were kept separately from participants' personal details in order to ensure anonymity and confidentiality. Only the researcher and principal investigator had access to the data as well as the document linking the identity of the participant and the data. All participants were given or asked to create a unique identifier in order to anonymise and facilitate withdrawal of data. Participants were told of their right to withdraw their data up until the point of data analysis. After which any personal data (as defined by the Data Protection Act, 1998) would be destroyed, and only anonymised raw data and summary data would be retained after that. Participants were informed that after the research has been completed and provided that they do not withdraw their raw data, summaries of non-personal data (demographics, responses in questionnaires and performance test results) may be retained long-term as part of a larger data set and used for publication. At the end of each study, participants were debriefed about the purpose of the study and the researcher's contact details were also given, if they wished to withdraw their data.

Chapter 3 – Examining the dual driver advantage hypothesis

Chapter overview

Literature reviewed in the first chapter highlighted that car drivers are susceptible to overlooking motorcycles on the road (e.g., ACEM, 2009; Crundall et al., 2012) and indicated that dual drivers are less likely to collide with motorcycles than car drivers (Magazzù et al., 2006). There was some evidence hinting at a familiarity-based perceptual advantage and a gap in the literature regarding the possibility of a motorcycle-specific perceptual advantage for dual drivers was identified. As dual drivers are more familiar with motorcycles than car drivers, they may have developed better perceptual skills for spotting motorcycles. In turn, the motorcycle-specific perceptual advantage may protect dual drivers against motorcycle collisions. Therefore, the aim of this chapter is to determine whether dual drivers are better at detecting motorcycles than car drivers, with a focus on looking and processing.

Three experiments will be reported in this chapter, which will examine hazard prediction ability, visual search and processing thresholds for motorcycles. The first experiment compares dual drivers' and car drivers' accuracy in predicting motorcycle hazards before they fully develop. This could inform whether dual drivers are better at predicting hazardous situations involving motorcycles, which could in turn account for their low risk of colliding with motorcycles. The second experiment compares detection times for motorcycles between dual drivers and car drivers in a visual search task. The results can further our understanding on whether dual drivers have an enhanced attentional set for motorcycles during hazard search compared to car drivers. If dual drivers were faster to detect motorcycles in an array than car drivers, familiarity could have increased sensitivity to the motorcycle features and shapes such that they stand out to dual drivers. As a result, dual drivers may have a lower likelihood of colliding with motorcycles than car drivers because they can spot the motorcycle quickly. Lastly, the third experiment investigates whether dual drivers have a lower processing threshold for motorcycles compared to car drivers in a T-junction task. This experiment is designed to test drivers' accuracy in detecting the presence of a vehicle and identifying a motorcycle or car when presented at short durations. The results can help determine whether there are any differences in processing thresholds for motorcycles between dual drivers and car drivers. A lower processing threshold for motorcycles would mean that dual drivers require less visual information to detect or identify a motorcycle than car drivers, and thus avoid a collision with the motorcycle in time.

3.1 Experiment 1: Are dual drivers better at predicting motorcycle hazards than car drivers?

3.1.1 Introduction

In Chapter 1, the failure to perceive an approaching motorcycle was identified as the primary cause for car-motorcycle collisions, with the car driver more frequently at fault than the motorcyclist. In order to successfully avoid a collision, drivers need to be able to detect potential hazards and appraise the level of risk, before making an appropriate behavioural response. The ability to detect potential hazards is therefore the first criteria to hazard avoidance. To answer the research question of whether dual drivers have a motorcycle-specific perceptual advantage, the first experiment aims to assess dual drivers' and car drivers' ability to perceive motorcycle hazards. This ability to perceive and be aware of potential hazards is known as hazard perception, which is also defined as the ability to predict or anticipate potentially hazardous situations on the road (Crundall, 2016; Horswill, 2016; Horswill & McKenna, 2004). Hazard perception ability is a driving skill that has been consistently shown to correlate with collision involvement (Horswill & McKenna, 2004; Wells et al., 2008) as well as the amount of driving experience (Horswill et al., 2008; Wallis & Horswill, 2007). For instance, novice drivers with little on-road driving experience have impoverished hazard perception abilities than experienced drivers even after accounting for age and risk-taking tendencies (Scialfa et al., 2011). Therefore, comparing hazard perception abilities for motorcycle hazards could inform whether dual drivers have a perceptual advantage that translates to low collision likelihood with motorcycles.

There is some evidence that dual drivers have better hazard perception abilities than car drivers when driving (Horswill & Helman, 2003; Rosenbloom et al., 2011). Horswill and Helman (2003) compared the driving behaviours and hazard perception abilities of dual drivers and car drivers in a driving simulator. They found little difference between driver groups in driving behaviours, such as speed and following distance, but dual drivers exhibited better hazard perception abilities than car drivers. In addition, another study found that dual drivers were faster at responding to hazards that gradually unfold than car drivers, but not for abrupt hazards (Underwood et al., 2013). The detection of gradual hazards relies on accurate interpretations of hazard precursors, which are cues in the environment that indicate what is about to happen. Therefore, the findings suggest that dual drivers are better at identifying and interpreting these hazard precursors than car drivers. This is possible because compared to car drivers, dual drivers have a more diverse experience of on-road hazards as riders and drivers, thus resulting in a better cognitive schema for hazard searching and looking (see Pradhan & Crundall, 2016). In contrast, abrupt hazards capture attention from the bottom-up and thus would attract similarly fast responses from dual drivers and car drivers (Underwood et al., 2013).

However, these studies report response times as a measure of hazard perception ability, as with conventional hazard perception tests, where faster response times indicate better hazard perception abilities (Crundall, 2016; Horswill & McKenna, 2004). In traditional hazard perception tests, drivers watch a series of clips from the perspective of the driver and are tasked to make a button response as soon as they detect a developing hazard. Drivers score more points for earlier responses, although any responses outside a temporal scoring window would be considered incorrect. The scoring window represents the period of time that drivers have to avoid the hazard in time. The problem with using response times as an indicator of hazard perception ability is that response times could be influenced by the driver's appraisal of how hazardous or risky the situation is. There is evidence suggesting that response times in hazard perception tests are confounded by the driver's threshold of acceptable risk or response criterion (Borowsky & Oron-Gilad, 2013; Crundall, 2016; Pradhan & Crundall, 2016). Borowsky and Oron-Gilad (2013) conducted a standard hazard perception test and analysed drivers' reasons for responding to a hazard. The authors found that responses in the hazard perception test were based on the driver's perceived likelihood of a crash. This means that a driver who detected the hazard early on but has a high threshold for risk acceptance would score similarly on the traditional hazard perception test as another driver who was slow to detect the hazard. The influence of risk acceptance on response times could potentially account for several studies that previously failed to find an advantage for experienced car-only drivers compared to novice car-only drivers (Borowsky et al., 2010; Sagberg & Bjørnskau, 2006).

A better approach for assessing hazard perception ability without the confound of risk acceptance is to measure accuracy on a hazard prediction test. The hazard prediction test involves showing clips taken from the perspective of the driver and immediately prior to the onset of the hazard, the screen is occluded. Drivers are then tasked to predict what happens next and four multiple choice options are provided (see Ventsislavova & Crundall, 2018). The hazard prediction test is thus designed to measure drivers' situational awareness during hazard search and the ability to interpret the relevant hazard precursors, which more accurately assesses hazard perception ability than previous hazard perception tests (Crundall, 2016). Previous research has shown that the hazard prediction test can reliably differentiate experienced from novice drivers across cultures (Ventsislavova et al., 2016). The possibility of a dual driver advantage in motorcycle hazard perception ability as compared to car drivers can therefore be examined using the hazard prediction test.

Current study

The current study investigates the hypothesis that dual drivers have superior motorcycle hazard perception ability using a hazard prediction test. Previously, research into dual drivers' hazard perception abilities only examined overall hazard perception ability and did not differentiate between motorcycle hazards and non-motorcycle hazards. Aside from the limitation of response times as a measure of hazard perception ability, the current literature does not tell us whether dual drivers have a motorcycle-specific benefit for hazard perception. Therefore, to address this gap in the literature, the current study investigates whether dual drivers are better at predicting motorcycle hazards than car drivers using a hazard prediction test. In turn, this can provide insight into why dual drivers have a lower likelihood of colliding with motorcycles than car drivers (i.e., Maggazù et al., 2006). A comparison between dual drivers' and car drivers' ability to predict motorcycle and non-motorcycle hazards could be made to test whether a motorcycle-specific hazard perception advantage exists for dual drivers. The hazard prediction test used in the current study was adapted from Goodge et al. (submitted) and included two motorcycle hazards out of a total of ten hazards. In their study, the motorcycle clips were originally cut at the point where the conflicting motorcycle began to appear. However, at this occlusion point, the limited amount of visual information may be insufficient visual information for drivers to process and recognise the motorcycle (Crundall, Clarke et al., 2008). Therefore, the motorcycle clips in the current study were cut later such that the motorcycle was briefly visible to allow drivers to process it. Further details are outlined in the methods section. If dual drivers have an advantage for looking and processing motorcycles, they should be better at predicting motorcycle hazards in the hazard prediction test than car drivers.

Usually, the clips showed in the hazard prediction test and hazard perception test are presented on single screens. However, during the test, participants' field of view are restricted by the single screen and information that would normally be available from side mirrors and windows are absent (Alberti et al., 2014). Alberti et al. (2014) compared the effect of wide or narrow field of view on experienced and novice drivers' horizontal spread of fixations during a simulated drive, using three screens or single screen respectively to present hazard perception clips. They found that experienced drivers but not novice drivers increased their spread of visual search in the condition with three screens, suggesting that a wide field of view better encapsulates experiential differences in drivers as compared to a narrow field of view. In light of this, the current study presented the clips in the hazard prediction test in a 360° virtual environment on virtual reality headsets. The 360° environment allows participants to have a wider field of view and a greater range of head and eye movements that resembles real world driving more closely than on single screens (see Underwood, Crundall, & Chapman, 2011; Goodge et al., submitted). This is important because in Crundall, Clarke et al.'s (2008) framework, the first stage for understanding motorcycle collisions is whether drivers looked. Therefore, examining dual drivers' hazard perception ability can also provide some insight into whether they are more likely to know where to look during hazard search, that allows them to better predict motorcycle hazards than car drivers. This can be inferred by accuracy on the hazard prediction test because participants need to have looked at the appropriate locations in order to predict the hazard accurately.

The aim of the current study is to investigate whether dual drivers are better at predicting motorcycle hazards than car drivers in a virtual reality hazard prediction test. Based on previous literature suggesting that dual drivers have better overall hazard perception ability than car drivers, it is hypothesised that dual drivers would be more accurate at predicting motorcycle hazards than car drivers.

3.1.2 Methods

Design

A 2x2 mixed design was used. The between-participants variable was driver group, which compared dual drivers and car drivers. The within-participants variable was the type of hazard that drivers had to predict in the clip (motorcycle hazard vs non-motorcycle hazard). The dependent variable was accuracy – whether their answer in the multiple-choice test was correct or incorrect. There were two motorcycle clips and eight non-motorcycle clips in this analysis.

Participants

Sixty-five participants took part in this study, but two participants did not complete the study due to motion sickness. One motorcyclist did not have a driving license and was also excluded from the analysis. The final sample consisted of 29 car drivers (9 females, 20 males) and 33 dual drivers (4 females, 29 males). Dual drivers were recruited by reaching out to local motorcycle riding clubs and attending a motorcycle exhibition show in Birmingham. Car drivers were recruited via opportunistic sampling in Nottingham and London. The demographics of each driver group are provided in Table 3.1. Driving experience and hours driven per week were significantly different between driver groups and were thus included as covariates in the analysis. Participants were not screened for visual skills such as visual acuity or contrast sensitivity.

Table 3.1: Demographics, driving history and habits of each driver group

	Car drivers	Dual drivers	*p*
Age	38.3 years (10.0)	43.2 years (12.0)	.083
Driving:			
Driving experience	18.5 years (9.4)	23.7 years (10.9)	.048*
Annual mileage	9051.7 miles (6700.7)	11201.52 miles (8677.6)	.28
Hours driven per week	7.8 hours (4.4)	13.5 hours (14.2) [a]	.045*
Riding:			
Riding experience	-	14.1 years (12.1)	-
Annual mileage	-	4253.12 miles (5442.1)	-
Hours ridden per week	-	7.9 hours (10.8)	-

[a] one missing case

*Below significance threshold of .05

Materials and stimuli

A road safety clip from Road Safety Scotland (2017) designed for virtual reality was taken from YouTube and was shown before the test to acclimatise participants with the 360° environment. For the experimental task, ten 360° CGI clips depicting a continuous drive were used. There was one hazard in each of the clips and two out of the ten clips contained motorcycle hazards. The clips were cut immediately before the hazard became hazardous and the screen would turn black, at which point the multiple-choice question 'What happens next?' would be displayed. There were four choices, of which only one was the correct answer. Each clip followed on from the previous so that participants could see what actually happens next. The clips were obtained with permission from an ongoing project at the Transport Research in Psychology group at Nottingham Trent University. Therefore, the point at which the non-motorcycle clips were cut and the multiple-choice options had been previously determined by the researchers. Further information about the development of the clips can be obtained from Goodge et al. (submitted).

However, several changes were made to the motorcycle clips to better answer the current research question. First, the occlusion points for the two motorcycle clips were delayed four frames (~160ms) later than the original motorcycle clips. This was to ensure that the motorcycle was visible for a longer time than the original clips for processing to occur. Figures 3.1 and 3.2 show screenshots of the two motorcycle clips before they were cut in the current study and Figure 3.3 shows an example of a non-motorcycle hazard clip before it was cut. A brief summary of the remaining clips is provided in Appendix A. Secondly, for both motorcycle clips, one of the multiple-choice options was changed to reflect similar hazards developing in the same part of the screen. The correct option in both motorcycle clips was 'An oncoming motorcycle prevents you from turning'. In the first motorcycle clip, one of decoy options was changed from 'The HGV decides not to turn right and proceeds straight across the junction, narrowly missing you' to 'A car appears from behind the HGV and forces you to give way'. This option was changed because at the later occlusion point, participants would have seen the HGV come to a complete stop at the junction. In the second motorcycle clip, one of the decoy options was changed from 'One of the cars waiting in the oncoming lane closes the gap into the side road, preventing you from turning' to 'One of the cars waiting in the oncoming lane cuts into your lane, forcing you to stop'. This prevents participants from picking the correct option simply because they detected movement from the motorcycle appearing from behind the car in the oncoming lane (see Figure 3.2).

The clips were presented using Oculus Go headsets and the Oculus Go headsets were screen cast onto a laptop to allow the experimenter to confirm that the task was running smoothly.

A Qualtrics form was created to record participants' answers and their demographics regarding their age, sex, driving (and riding if applicable) experience, annual (riding) mileage and number of hours a week they spend driving (riding).

Figure 3.1: Screenshot of first motorcycle clip when it was cut; motorcycle appears from behind the HGV (red circle not in the actual test)

Figure 3.2: Screenshot of second motorcycle clip when it was cut; motorcycle appears from behind the black car (red circle not in the actual test)

Figure 3.3: Screenshot of general hazard clip when it was cut; pedestrian with buggy appears between parked vehicles on the left (red circle not in the actual test)

Procedure

All participants gave informed consent prior to the start of the study. They were informed about the potential side effects of motion sickness within virtual environments and their right to stop at any point during the study. Refreshments and sweets were available in case participants felt unwell. Once the headset was in place, participants watched the Road Safety Scotland (2017) clip as a 'test drive' and were given as much time as they needed to adjust to the virtual environment. The first clip of the hazard prediction test was played after participants felt comfortable enough during the 'test drive'. Participants responded verbally by reporting aloud the number of the option chosen to the researcher and the researcher would enter their response into Qualtrics. There was no time limit to respond. At the end of the test, participants were asked to fill in the demographics questionnaire, and were then thanked and debriefed.

Statistical analysis

Data were analysed using the statistical programming language R (R Core Team, 2020). The original analysis was a generalised linear mixed model with a binomial link function to model response accuracy, with between-participant and between-stimulus random effects. Driving experience and hours driven per week were included as covariates. However, the inclusion of random effects did not improve the model for this analysis. Comparison of model performance showed that the full model with the specified random structure did not perform better than the traditional general linear model without random effects (see Table 3.2). The AIC (Akaike information criterion) values are also reported in the table, with lower values representing a better model fit of the data (Akaike, 1974). Therefore, for parsimony, the final model used in the analysis was without random effects:

Accuracy ~ Driver group + Clip type + Driver group:Clip type + Driving experience + Hours driven per week

Reference/Baseline categories of fixed effects: Driver group - car driver, Clip type – Non-mc

The effects of driver group and hazard type were tested by dropping each effect in turn using a likelihood ratio test, which computes the change in model fit with and without the term of interest, as described in Chapter 2.

Table 3.2: Model fit comparison

Source	**AIC**	**χ^2 difference**	***p***
Full model	609.8	1.80	.62
GLM model without random effects	605.6		

3.1.3 Results

The mean proportion of correct responses for the motorcycle hazards and other general hazards is shown in Figure 3.4, followed by a breakdown of accuracy in each clip given in Figure 3.5. As participants were either correct or incorrect (0 or 1), mean proportion of correct responses ranged from 0 to 1. Hazard prediction was more accurate for the non-motorcycle clips than motorcycle clips (χ^2 (1) = 146.7, $p < .001$; *OR* = 0.069 [95% CI: 0.042, 0.11]). However, the effect of driver group was not significant (χ^2 (1) = 0.15, $p = .69$; *OR* = 0.92 [95% CI: 0.60, 1.41]) and the interaction between driver group and the type of hazard was also not significant (χ^2 (1) = 1.62, $p = .20$; *OR* = 1.87 [95% CI: 0.71, 5.05]).

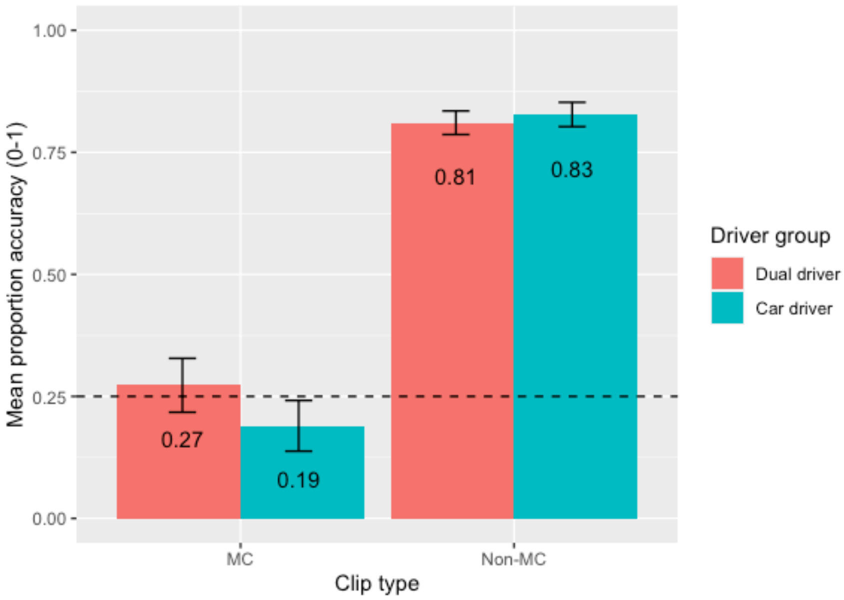

Figure 3.4: Mean proportion accuracy by driver group and type of hazard shown in the clip with +/-1 SE and chance performance at 0.25

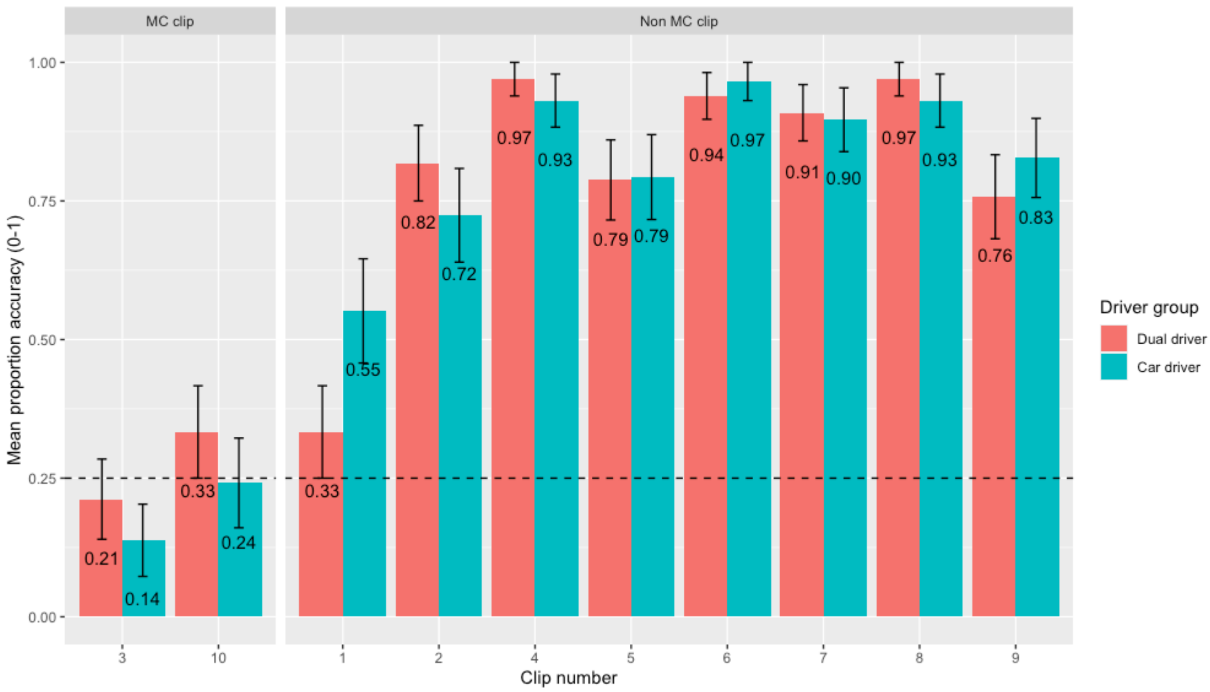

Figure 3.5: Mean proportion accuracy by driver group in each clip with +/-1 SE and chance performance at 0.25

3.1.4 Discussion

Using a virtual 360° hazard prediction test, the present study tested whether dual drivers were better able to predict motorcycle hazards than car drivers. This hypothesis was not supported – dual drivers were not more likely to accurately predict what happens next in motorcycle hazard clips than car drivers. Furthermore, there was no difference between driver groups in predicting non-motorcycle hazards. This is the first study to examine motorcycle-specific hazard perception abilities in dual drivers and the findings provide initial evidence against a dual driver advantage for predicting motorcycle hazards. Although dual drivers are likely to be more familiar and knowledgeable of potential motorcycle hazards than car drivers, the findings suggest that prior knowledge does not translate to increased situational awareness for motorcycle hazards in a hazard prediction test. Therefore, the present findings do not support the hypothesis that dual drivers are less likely to collide with motorcycles because they have better hazard perception for motorcycles than car drivers.

An implication of the lack of dual driver advantage in predicting motorcycle hazards is that dual drivers were not more likely than car drivers to have looked at and processed the motorcycle during hazard search. This is because the design of the hazard prediction test requires participants to have looked in the right locations at the right time and processed it in time to predict the hazard accurately. Therefore, it can be inferred from participants' response accuracy whether they looked at and processed the motorcycle in time. However, it should be noted that due to practical limitations, eye movement data were not collected in this study, which could have helped establish whether participants fixated on the motorcycle. Although a virtual reality headset with built-in eye trackers could have been acquired for this study, the sensors needed to track the participants' head and eye movements in three-dimensional space could not be set up given the limitation of the venue where testing occurred (Birmingham National Exhibition Centre). Another potential limitation of the study is that only two out of the ten clips were motorcycle hazard clips and two critical trials may have been insufficient to detect a difference in driver groups. If a dual driver advantage for predicting motorcycle hazards did exist, the current study design had no more than 0.4 power to detect an effect size of 0.4, assuming a similar effect size to previous research (Horswill & Helman, 2003). This was a practical limitation as the clips were developed for another research project and additional motorcycle clips could not be included in the hazard prediction test. However, an argument could be made for keeping the proportion of motorcycle clips low to replicate the low prevalence of motorcycles in real life.

The lack of driver group difference in predicting non-motorcycle hazards in the current study contradicts previous studies on dual drivers' hazard perception abilities, which reported that dual drivers have superior hazard perception abilities in general (e.g., Horswill & Helman, 2003; Rosenbloom et al., 2011; Underwood et al., 2013). This could be attributed to the different ways in which hazard perception ability was measured in the current study and previous studies. The current study measured accuracy in predicting hazards before they occurred, whereas previous studies measured response times to hazard detection. This suggests that while dual drivers may not be better at predicting hazards than car drivers, they may respond quicker to hazards. As mentioned previously, differences in response times to hazard detection may reflect differences in risk acceptance and/or actual differences in hazard detection (e.g., Borowsky & Oron-Gilad, 2013). Theoretically, it is possible that dual drivers are quicker to detect motorcycles than car drivers during hazard search, even after accounting for variations in risk acceptance. This is because dual drivers are likely to be more familiar with motorcycles and their features than car drivers, and their attentional set may be more tuned in to motorcycles. An attentional set helps identify targets by biasing exogenous attention to objects that match the attentional set (Reeder & Peelen, 2013). As a result, motorcycles may stand out to dual drivers, such that motorcycles are processed quickly during hazard search, allowing quicker detection times as compared to car drivers. To examine whether dual drivers are faster to detect motorcycles than car drivers without implicating drivers' risk acceptance, a visual search task could be used where the objective is to detect motorcycles rather than hazards. Therefore, the following study compares dual drivers' and car drivers' search times for motorcycles embedded within traffic scenes.

To summarise, dual drivers did not show an advantage for predicting motorcycle hazards. This suggests that dual drivers do not have a lower likelihood of colliding with motorcycles because they were better able to anticipate a developing motorcycle hazard and thus avoid a collision in comparison to car drivers. Having found that dual drivers were not better at predicting motorcycle hazards than car drivers, the following study focuses on whether dual drivers are quicker at searching for motorcycles. The hypothesis that dual drivers have an advantage at detecting motorcycles during visual search due to attentional capture from motorcycles is tested, by comparing how quickly dual drivers and car drivers can detect motorcycles in an artificial search array.

3.2 Experiment 2: Are dual drivers quicker at spotting motorcycles than car drivers?

3.2.1 Introduction

Visual search is relevant to understand how drivers perceive hazards because they need to be constantly scanning for hazards and processing the scene. However, visual search in driving is more complex than typical visual search paradigms in laboratory studies. Unlike laboratory conditions, drivers do not have a preview of the target they are searching for (Hollingworth, 2012). Furthermore, search targets in driving, like in real-world search, are often defined categorically rather than pre-specified targets (see Schmidt & Zelinsky, 2009). This means that target detection is dependent on knowledge and representations that define the target category. Categorical search is therefore aided by features that are representative of the category (Hollingworth, 2012). Supporting evidence comes from research indicating that target detection increases for objects that are more prototypical of the category (Maxfield et al., 2014; Robbins & Hout, 2015). Based on these reasons, it is logical to expect that prior knowledge of the target category plays an important role in categorical search.

During categorical search, top-down knowledge of features that help identify the target in turn biases attention to objects that have these features from the bottom-up. Categorical search thus requires a combination of top-down and bottom-up attentional guidance. Reeder and Peelen (2013) proposed that prior knowledge activates an attentional set that is used to bias exogenous attention during search, which is commonly referred to as a search template. According to the authors, visual search within natural scenes is akin to matching visual input to the search template. Particularly in driving, the lack of specific targets and target previews means that the search template is defined by features that are representative of the target category, rather than a specific visual image of the target (Yang & Zelinsky, 2009). The search template biases exogenous attention towards diagnostic features such that objects matching these features stand out and capture attention more readily, thus aiding detection (see also Reeder et al., 2015). This occurs through early visual input feeding forward to higher-level visual processing and a feedback process of long-term memory representations to increase bottom-up activation of features that match the properties of the target category, resulting in a priming effect of representative features (see Wolfe, 1994, 2014). There is also neuroimaging evidence illustrating the bias in exogenous attention; neuronal responses to objects with features matching the search template is heightened during search, even when it is not currently attended to (Saenz et al., 2002; Serences & Boynton, 2007). The attentional set activated during categorical search may therefore vary according to individual differences in top-down knowledge and memory representations of the target category.

Current study

As the search template relies on prior knowledge of the target category, an implication is that robust long-term memory representations may enhance the effectiveness of the search template (Olivers, 2011). Given that dual drivers are more familiar with motorcycles than the average car driver, they should have more robust long-term memory representations of motorcycles, which would in turn facilitate visual search for motorcycles (see also Christie & Klein, 1995). There is some evidence suggesting that familiarity – in terms of natural expertise rather than experimentally familiarised stimuli – guides attention during categorical search. Hershler and Hochstein (2009) found that car experts were quicker to detect cars in an array of distractor objects than bird experts and vice versa for bird stimuli. Similarly, Golan et al. (2013) found that objects from experts' domain of expertise were detected faster and more efficiently, as measured by shallower search slopes with increasing number of distractors, than objects from another category. In addition, as categorical search is influenced by prior knowledge of features that are representative of the target category (Hollingworth, 2012), dual drivers may also benefit from having representations of prototypical features of motorcycles during visual search. This is supported by Robbins and Hout's study (2015), which found that typical members of a cued category were detected more quickly than atypical members.

However, the extent to which familiarity facilitates real world categorical search is still uncertain as a recent review noted that previous research is largely conducted on overlearned stimuli such as familiar faces and words rather than semantic categories (Wolfe & Horowitz, 2017; see also Nako, Wu, & Eimer, 2014). The familiarity effect was reportedly attenuated for categories that are not overlearned or are less well-defined such as clothing, as compared to overlearned stimuli such as alphabets and numerals (Nako, Wu, Smith, & Eimer, 2014). The recorded event-related potential component, which indicates the activation of search template, was smaller and slower [activation] during a categorical search task for clothes and kitchen objects than for overlearned stimuli like letters (Nako, Wu, Smith, & Eimer, 2014). In summary, while some research suggests that familiarity may improve detection in a search task, there is currently insufficient evidence on whether the benefit of familiarity is robust. Therefore, the current study investigates the hypothesis that familiarity with motorcycles leads to quicker motorcycle detection using a categorical search task. Quicker motorcycle detection rates from dual drivers could provide some insight into whether dual drivers have a perceptual advantage for motorcycles as compared to car drivers.

Previous studies investigating visual search for motorcycles often use an absent-present design whereby each traffic scene is presented one at a time, and participants make a button response if a motorcycle is present (Gershon et al., 2012; Sanocki et al., 2015). As the current study examines whether dual drivers have an enhanced attentional set for motorcycles, the design used by previous studies may not be suitable for comparing detection times between driver groups. This is because search for a motorcycle within a traffic scene may be facilitated by location-based cues and cognitive schemas. Dual drivers may simply pay attention to locations where motorcycles tend to occupy, such as regions closer to the centre line or kerb. Therefore, to assess whether dual drivers are quicker to detect motorcycles than car drivers due to attentional capture, it is important to reduce location-based influences in the current study. A novel design is used to test the hypothesis that familiarity with motorcycles increases sensitivity to motorcycle shapes and features during search. Multiple traffic scenes will be presented in a visual search array, with only one of the traffic scenes containing a motorcycle. Participants are therefore searching for the image that contains a motorcycle rather than a motorcycle within one traffic scene. This design minimises the influence of location-based cues and experienced-based guidance during search because the target may appear anywhere in the array. By varying the location of the motorcycle target in the array, participants have to rely on their attentional set to search for the motorcycle. A control condition is also included to compare baseline search times using a target that is equally unfamiliar to both driver groups – hackney carriages. Hackney carriages were chosen as the control task because they involve serial search and should be equally difficult to detect for all participants. It is expected that hackney carriages would be more difficult to detect than motorcycles in the search array, therefore the difference in response times between the two target vehicles are not of interest in the present study. Instead, the research question focuses on the interaction between driver group and target vehicle, in particular whether there is a difference between dual drivers and car drivers in the motorcycle condition but not hackney carriage condition.

The aim of the current study is to determine whether dual drivers' familiarity with motorcycles facilitates detection of motorcycles in an artificial array of traffic scenes. It is hypothesised that dual drivers would be quicker to detect motorcycles than car drivers, because motorcycle features are more prominent in their attentional set during search. There would be no difference between driver groups in detecting hackney carriages.

3.2.2 Methodology

Design

A 2x2 mixed design was used. The between-participants variable was driver group, which compared dual drivers and car drivers. The within-participants variable was type of search target – motorcycle or hackney carriage (control). The outcome variable was time taken to find the target vehicle.

Participants

Forty-nine participants took part in this study, of whom 24 were dual drivers. One dual driver and two car drivers did not provide demographics and were thus excluded from the analysis. The final sample consisted of 23 car drivers (11 females, 12 males) and 23 dual drivers (6 females, 17 males). Dual drivers were recruited by reaching out to motorcycle accessory shops in Nottingham and attending Bike Nights held by local motorcycle riding clubs. Car drivers were recruited from local sports clubs in Nottingham and Nottingham Trent university staff. The demographics of each driver group is shown in Table 3.3. Age was significantly different between driver groups and was included as a covariate in the analysis. Participants were not screened for visual skills.

Table 3.3: Demographics of each driver group

	Car drivers	**Dual drivers**	***p***
Age	34.96 years (9.5)	41.87 years (10.01)	0.020*
Driving experience	17.04 years (8.97)	21.5 years (11.11)	0.13
Riding experience	-	15.21 years (13.05)	-

*Below significance threshold of .05

Materials and stimuli

A total of 320 coloured images of traffic were taken from the internet and cropped such that at least one vehicle in the image took up roughly a third of the image size. Twenty arrays were created with 16 unique images in each array arranged in a 4x4 grid (see Figure 3.6 for an example array and refer to Appendix B for all arrays). In every array, there was one image containing a motorcycle within the traffic and one image containing a hackney carriage within the traffic. This ensured that all participants were subject to the same distractor images. The location of the target vehicles was assigned randomly across arrays but remained the same across participants. All participants saw the same 20 arrays once and were instructed to search for a motorcycle for ten of the arrays and to search for a hackney carriage for the other ten arrays.

Figure 3.6: Example array of 16 images, of which one contains a hackney carriage and one contains a motorcycle

The study was coded on Unity and run in full screen mode on a laptop with a screen size of 13.3" and a display resolution of 1366 pixels x 768 pixels (aspect ratio 16:9). The resulting array measured approximately 20 cm x 16 cm and each image within the array measured approximately 5 cm x 4 cm. When participants were sat at a distance of approximately 60 cm, the screen subtended 29.2° along the horizontal axis and 21.0° in the vertical axis. A timer was used to record the time taken to find the target and was visible throughout the task. Participants could only advance to the next array if the correct image was found. If the wrong image was selected, the border of the image would turn red to prompt participants to continue searching. A short demographics questionnaire was created to asked participants to enter their age, sex, driving (and riding) experience at the end of the 20 arrays.

Procedure

Before the study began, an information sheet and consent form were provided, and participants were given a chance to clarify any questions as there were no practice given. As the entire study took less than ten minutes, participants were not compensated. Participants responded by left-clicking on the image with a mouse. Instructions were provided prior to each of the tasks, to inform participants what type of vehicle they had to search for. The order of the two search tasks was generated randomly using a simple random generator within Unity; resulting in 21 participants who searched for motorcycles first (10 experienced drivers and 11 dual drivers) and 28 participants who searched for hackney carriages first (15 experienced drivers and 13 dual drivers). Within each task block, the arrays were presented in the same order. For the hackney carriage task, the instructions provided clarification in case some participants misunderstood hackney carriages to include minivans, which are increasingly being used as private hires (see Figure 3.7). Upon completing the visual search task, participants were asked to fill in the demographics questionnaire, before being thanked and debriefed.

Figure 3.7: Instructions for the hackney carriage task

Analysis

Data were analysed using the statistical programming language R (R Core Team, 2020). Four trials were more than three standard deviations from the mean and were identified as outliers, however exclusion of these data points did not change the results therefore all 920 trials were analysed (refer to Appendix C for the analysis without outliers). A linear mixed effect model was fitted using the lme4 package (Bates et al. 2015) and the analysis was conducted with likelihood ratio tests as described in Chapter 2. The full model with the specified random structure performed better than the traditional general linear model without random effects (see Table 3.4) and is as follows:

Time taken ~ Driver group + Target + Driver group:Target + Age + (1|Participant) + (Target|Array)

Reference/Baseline categories of fixed effects: Driver group - car driver, Target – hackney carriage

Table 3.4: Model fit comparison (below significance threshold of .05)*

Source	AIC	χ^2 difference	*p*

Full model	1723.2	491.51	< .001*
GLM model without random effects	2200.7		

3.2.3 Results

The mean time taken to detect the target vehicles according to driver group is shown in Figure 3.8.

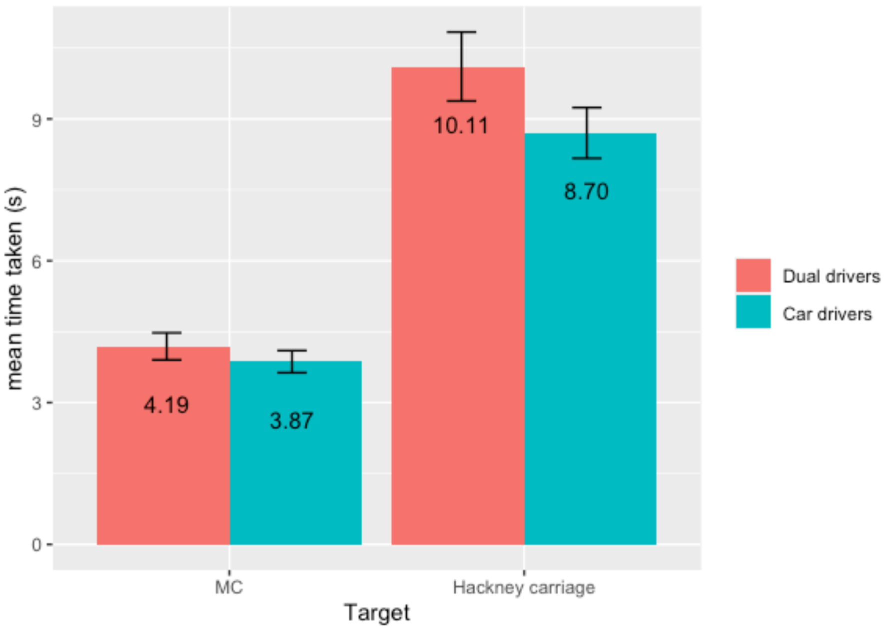

Figure 3.8: Mean time taken with +1/-1 SE to locate motorcycles or hackney carriages by driver group

As the assumption of normality of the residuals was violated (as indicated by a Q-Q plot), a logarithm transformation was performed on time taken. Comparing two models with and without the fixed effect driver group, a likelihood ratio test revealed that driver group was not a significant factor (χ^2 (1) = 0.11, p = .74; β = -0.021 [95% CI: -0.15, 0.10]). A likelihood ratio test for target found that time taken differed significantly according to type of target vehicle (χ^2 (1) = 27.20, p < .001). Across both driver groups, participants took significantly longer to search for hackney carriages (M = 9.41s, SD = 9.72) than motorcycles (M = 4.03s, SD = 3.96; β = -0.75 [95% CI: -0.94, -0.56]). Model comparisons with and without the interaction term found that the interaction between driver group and target was not significant (χ^2 (1) = 0.43, p = .51; β = -0.049 [95% CI: -0.20, 0.098]; see Figure 3.8 for means).

3.2.4 Discussion

The present study compared the search times for motorcycles in a categorical search task between dual drivers and car drivers, without the presence of location-based cues. It was hypothesised that dual drivers would be quicker to detect motorcycles in an artificial array of traffic scenes than car drivers because motorcycle features are likely to stand out to dual drivers, thus reducing search times. However, this hypothesis was not supported by the present findings. Dual drivers were not quicker to spot the motorcycle from the array of traffic scenes. The findings suggest that familiarity with motorcycles does not offer a domain-specific advantage for motorcycle detection during this particular instance of categorical search. Therefore, the notion that dual drivers have a perceptual advantage for motorcycles in terms of increased sensitivity to motorcycle features during hazard search is not supported.

Although research suggests that the attentional set activated during visual search is influenced by prior knowledge of the target category (Olivers, 2011; Reeder & Peelen, 2013), this was not observed in dual drivers in the present study. Dual drivers did not experience a stronger 'pop-out' effect from the motorcycle features and shapes than car drivers, despite being more familiar with motorcycles (see Reeder et al., 2015). One possible reason for the lack of dual driver advantage is that explicit instructions were given to search for motorcycles in the present study. The presence of instructions does not reflect usual hazard search conditions because the type of the hazard is usually not known. For example, drivers are usually searching for not only one category of hazards, such as motorcycles, but multiple categories, such as any approaching cars, motorcycles or pedestrians. The instructions likely prompted all participants to activate an attentional set that is solely focused on motorcycle features. As a result, car drivers may have benefitted from an attentional bias towards motorcycles, and this could have masked any familiarity-based advantage in motorcycle detection for dual drivers.

Therefore, it is possible that the similar search times observed in both driver groups could be due to task-specific demands, which negated any familiarity-based advantage in dual drivers. A similar effect was observed in another study, which found that experience with dieting and hence familiarity with healthy and unhealthy foods did not influence search efficiencies for familiar food categories (Wu et al., 2017). A greater activation in search template for familiar foods than unfamiliar foods, as measured by event-related potentials, was observed only when the food categories were irrelevant to the search task, suggesting that familiarity has a bigger effect on task-irrelevant than task-oriented attention capture. Together with the present findings, this implies that familiarity may have little influence on the effectiveness of a search template when actively searching for a familiar object but may evoke greater attentional capture when a familiar object is unexpectedly present.

Findings from the present study indicate that motorcycles are not more likely to capture dual drivers' attention during visual search compared to that of car drivers when participants are instructed to look for a motorcycle. A key difference between the current visual search task and previous studies is that the motorcycle could be located anywhere in the array (Gershon et al., 2012; Sanocki et al., 2015). As such, drivers in the present study could not rely on their cognitive schema of where motorcycles are likely to appear on the road to guide their visual search. In turn, drivers need to scan the entire array to look for the motorcycle and the time taken to detect a motorcycle may largely reflect scanning. As hazard perception consists not only of looking but also processing (Crundall, Clarke et al., 2008), there is still a need to examine drivers' visual processing abilities. Previous research suggests that experts require less visual information to identify objects of their expertise (e.g., Curby & Gauthier, 2009; Gayet et al., 2016). It is therefore possible that dual drivers' perceptual advantage arises in terms of lower processing thresholds for motorcycles than car drivers. A subsequent study using another perceptual task that focuses on drivers' processing thresholds for motorcycles is thus needed to substantiate the present findings. The study will also vary the type of vehicle presented across trials to prevent participants from anticipating a motorcycle and thereby masking any potential dual driver advantage, as highlighted above.

In conclusion, familiarity with motorcycles did not speed up search times of motorcycles in the present study. Dual drivers were not quicker to detect motorcycles from an artificial array of traffic scenes. The present findings suggest that a familiar attentional set for motorcycles is unlikely to have an observable benefit in terms of increased selective attention towards motorcycles when participants are instructed to search for a motorcycle. However, the visual search task using artificial arrays of traffic scenes places more emphasis on scanning than processing, which is also required for successful perception of the target. Previous research suggests that familiarity can facilitate perception by lowering processing thresholds for familiar objects. In light of this, the third experiment compares dual drivers' and car drivers' processing thresholds for detecting and identifying motorcycles in a T-junction task, where vehicles are presented at a highly predictable location to reduce the need for extensive scanning. The type of vehicle presented is also varied across trials to prevent participants from anticipating the presence of a motorcycle.

3.3 Experiment 3: Do dual drivers have a lower detection and identification threshold for motorcycles than car drivers?

3.3.1 Introduction

Object recognition is the process of bringing objects in the visual field into awareness and involves a sequence of detection, categorisation and identification (de la Rosa et al., 2011; Grill-Spector & Kanwisher, 2005). In object recognition literature, detection refers to judgements on the presence of an object, often at a superordinate level (e.g., animal, vehicle). Categorisation relies on recognition of the object's basic-level category, which refers to commonly used labels (e.g., bird, car, dog), whereas identification refers to recognition of its subordinate category (e.g., robin, sedan) or even the object's identity (e.g., my German Shepard). As basic-level categories are often the most accessible categories in everyday life, the terms categorisation and identification may be used interchangeably depending on the nature of the task. This is particularly so in driving because drivers only need to identify the basic-level category of the vehicle (e.g., car or motorcycle) but not the different models of cars and motorcycles (e.g., Volkswagen Golf or Kawasaki Ninja). For the purpose of clarity, identification rather than categorisation is used throughout the thesis to refer to identifying the type of vehicle. In driving, detection of a vehicle's presence alone is not adequate for hazard perception because drivers need to identify what the vehicle is, in order for subsequent appraisal and response to be made. For instance, identifying a vehicle as a motorcycle and not a car can help drivers better judge the gap between their vehicle and the motorcycle, given that motorcycles may appear further away than it actually is because of its small size (DeLucia, 1991). Therefore, detection and identification of the hazard is important for the hazard to be perceived. This is in line with Crundall, Clarke et al.'s (2008) framework for understanding motorcycle collisions, which outlines that motorcycles need to be perceived before appraisal can occur, and drivers must detect and identify the object as a motorcycle for successful perception.

In order to detect the presence of a vehicle and identify the type of vehicle present, the processing threshold of detection and identification must be reached accordingly. This means that enough visual information must be extracted and processed to reach the thresholds. The amount of information required for each threshold to be reached could in turn be determined by examining detection and identification accuracies when visual processing is disrupted. A common method is to vary the presentation times of stimuli. Using stimuli of common objects, de la Rosa et al. (2011) found that detection accuracy was higher at shorter presentation times than identification accuracy. This suggests that the threshold for detection was reached before identification.

Current study

The advantage of experts in object recognition observed in their domain of expertise is known as perceptual expertise or visual expertise (Harel, 2016; Harel et al., 2013). Through repeated exposure and experience, perceptual expertise develops in a bottom-up as well as top-down manner (Harel et al., 2013). As expertise is developed, bottom-up processing becomes more automatic and stimulus-driven. This is attributed to the increased selectivity of visual neurons and their stronger activation in response to objects of expertise. At the same time, higher-order functions such as conceptual knowledge and mental representations strengthen and become more accessible with expertise, thus facilitating object recognition from a top-down manner (Harel et al., 2013). Given that dual drivers are more familiar with motorcycles than car drivers in terms of exposure and experience, we can draw on the perceptual expertise literature when investigating whether dual drivers have a perceptual advantage for motorcycles.

The domain-specific nature of perceptual expertise suggests that an advantage for experts might be observed only for tasks that require category-specific knowledge. Recently, Crundall et al. (2017) used the T-junction task to compare the perceptual abilities of dual drivers and car drivers. Participants were tasked to imagine that they were emerging from a T-junction and checking for oncoming vehicles. An image of the road to the right of the T-junction was presented briefly and could either be empty or contain a car or motorcycle. Participants then responded whether a vehicle was present. Dual drivers were found to be more accurate at detecting the presence of vehicles than car drivers. However, when analysing responses according to the type of vehicle presented, there was no motorcycle-specific advantage for dual drivers. A potential reason could be that dual drivers did not benefit from their motorcycle expertise because detection only involves judging the presence of an object and does not require category-specific knowledge. Therefore, it would be interesting to disentangle detection from identification when examining whether dual drivers have superior perceptual abilities for motorcycles. The current study thus examines the hypothesis that dual drivers have lower processing thresholds for motorcycles using a T-junction task and extends previous T-junction tasks (e.g., Crundall et al., 2017; Crundall, Humphrey, & Clarke, 2008; Lee et al., 2015) by assessing both detection and identification accuracy. First, detection accuracy is measured by asking participants to judge whether a vehicle is present. Half of the trials will be empty roads and the other half will have a vehicle present. Identification accuracy is then measured by asking participants to identify the type of vehicle present. Out of the vehicle present stimuli, half will contain a motorcycle while the other half will contain a car. By measuring both detection and identification accuracy, the current study provides a thorough understanding of whether dual drivers' perceptual expertise for motorcycles might arise in terms of lower detection and/or identification thresholds.

As mentioned in the introduction, duration at which stimuli are presented could help investigate the processing thresholds of object recognition. Evidence suggests that experts have a lower processing threshold for objects of their expertise than non-experts. For example, Curby and Gauthier (2009) found that car experts were more accurate at making same-different discriminations of cars compared to non-experts at short presentation times. This suggests that experts are able to extract relatively more visual information than novices, thus allowing experts to make more accurate judgements, albeit at a subordinate level. Previously, Crundall et al. (2017) presented stimuli for 250ms to simulate typical fixation durations when looking in a direction to check for hazards. Besides the need for measuring identification accuracy, shorter presentation times may also be needed to distinguish between the detection thresholds of dual drivers and car drivers. Research suggests that information from the first 100ms of a fixation is sufficient to programme the next saccade (Ludwig et al., 2005), triggering a saccade shortly after (e.g., Crouzet et al., 2010; Kirchner & Thorpe, 2006). This implies that information extracted within the first 100ms can be useful in making some form of detection judgement. Therefore, the current study presents stimuli at 100ms and 250ms to investigate whether dual drivers are able to process more information than car drivers. It is hypothesised that vehicle detection and motorcycle identification accuracy would be higher for dual drivers than car drivers at 100ms. At 250ms, the difference in detection accuracy between dual drivers and car drivers may be attenuated but dual drivers may still exhibit an advantage at identifying motorcycles.

Similar to previous studies (e.g., Crundall et al., 2017; Crundall, Humphrey, & Clarke, 2008; Lee et al., 2015), the current study presented vehicles at one of three distances (near, mid, far). Previously, it was found that motorcycles were markedly difficult to detect at far away distances compared to cars presented at the same distance. Crucially, there was no dual driver advantage for detecting motorcycle stimuli at far distances (Crundall et al., 2017). However, a dual driver advantage may emerge for identifying faraway motorcycles at short presentation durations. Therefore, the current study also investigates the effect of challenging viewing conditions on dual drivers' ability to detect and identify motorcycles by examining factors such as presentation duration and vehicle distance.

The aim of the current study is to investigate the possibility of a dual driver advantage for motorcycle detection and identification thresholds. The type of vehicle, presentation duration, and distance at which vehicles are presented are varied to examine the influence of motorcycle familiarity under different viewing conditions. The hypotheses are:

Main effects-

1) Detection and identification accuracy will be lower for vehicles presented far away and for short durations.
2) Dual drivers will be more accurate at detecting and identifying vehicles across all distances and durations compared to car drivers.

Interactions-

1) Dual drivers will have an advantage in detecting and identifying motorcycles than car drivers. However, according to previous T-junction studies that found no motorcycle-specific advantage for dual drivers in detection, it is possible that the dual driver advantage may only be observed in identification accuracy.
2) The dual driver advantage for motorcycles will be enhanced under the most challenging conditions (i.e., far distance and short duration).

3.3.2 Methods

Design

A 2x2x2x2 mixed design was used. The between-participants variable was driver group, which compared dual drivers and car drivers. The first within-participants variable was vehicle type presented – motorcycle or cars. The second within-participants variable was presentation duration – 100ms or 250ms, and the last within-participants variable was distance of vehicle – mid or far. Trials with vehicles presented at near distance were excluded from the analyses because performance was close to ceiling for near trials (99.2% detection accuracy and 97.2% identification accuracy). There were two outcome variables – detection accuracy and identification accuracy.

Participants

Forty-four participants took part in this study, but two participants were excluded due to missing demographics data. The final sample consisted of 21 car drivers (7 females, 14 males) and 21 dual drivers (7 females, 14 males). Dual drivers and car drivers were recruited by advertising on social media. Participants were entered into a prize draw of £10 Amazon vouchers. The demographics and driving history of each driver group are shown in Table 3.5. As age was significantly different between driver groups, age was included as a covariate in the analysis. Participants were not screened for visual skills.

Table 3.5: Demographics and driving history of each driver group

	Car drivers	Dual drivers	p
Age	31.4 years (8.9)	39.7 years (13.4)	0.024*
Driving experience	13.0 years (7.6)	19.3 years (12.4)	0.059
Annual mileage	7524 miles (4214)	10958 miles (12783)	0.254
Riding experience	-	11.5 years (13.1)	-

* Below significance threshold of .05

Materials and stimuli

Ten digital colour photographs of T-junctions in Nottinghamshire were taken from the perspective of a driver waiting to pull out from a side junction. Photographs of cars and motorcycles were also taken around Nottinghamshire. In each T-junction, a car and a motorcycle were cropped out from their original photographs and edited into the T-junction photograph at each of the three distances – near, mid or far. This produced three car and three motorcycle stimuli for each T-junction. At each distance, the location at which vehicles were placed were determined by estimating where they best fit the perspective of the road. The cars and motorcycles were presented at the same locations in each condition, and they were 2 cm, 1 cm and 0.5cm in height on the screen at near, mid and far respectively. At near, mid and far respectively, the resulting stimuli for cars were 0.54%, 0.13% and 0.034% of the overall image, and for motorcycles were 0.22%, 0.056% and 0.014% of the overall image. To prevent participants from anticipating a vehicle in the stimuli, an equal number of empty junctions as vehicle stimuli were presented. All the stimuli were shown twice, once at 105.88ms (henceforth 100ms) and once at 247.059ms (henceforth 250ms). Participants therefore completed 240 trials. A demographics questionnaire was created to collect information regarding participants' age, sex, driving (riding) experience and annual driving (riding) mileage.

The study was programmed with E-prime 3 and presented on a CRT monitor with a refresh rate of 75 Hz. Stimuli were presented in full screen at 1024 X 768 resolution on a 17.2" monitor with 4:3 aspect ratio. When participants were sat at a distance of approximately 45 cm, the screen subtended 42.5° along the horizontal axis and 33.4° in the vertical axis.

Procedure

Participants gave informed consent prior to the experiment and were instructed to imagine that they were checking if it was safe to pull out from a T-junction, and to report only oncoming vehicles (i.e., not parked vehicles or vehicles facing away from them). A fixation cross was shown for one second before each stimulus to prompt participants to re-centre their gaze to the middle of the screen. After each image, participants responded to two questions on the screen by left-clicking on the options shown on the screen with the mouse. The first question was a simple forced-choice yes-no question to determine if they had seen any vehicles (i.e., detection). The second question was a multiple-choice question on the type of vehicle they had seen. The options were 'car', 'motorcycle', 'not applicable' if participants thought it was an empty road, or 'don't know'. For each of the detection and identification questions, participants had to rate their confidence in their responses on a scale of 1-10. Participants went through a practice block of six trials with stimuli not used in the actual study. Feedback was not provided during the experiment. Figure 3.9 summarises the experimental procedure. At the end of the experiment, participants were asked to complete the demographics questionnaire. They were then thanked and debriefed.

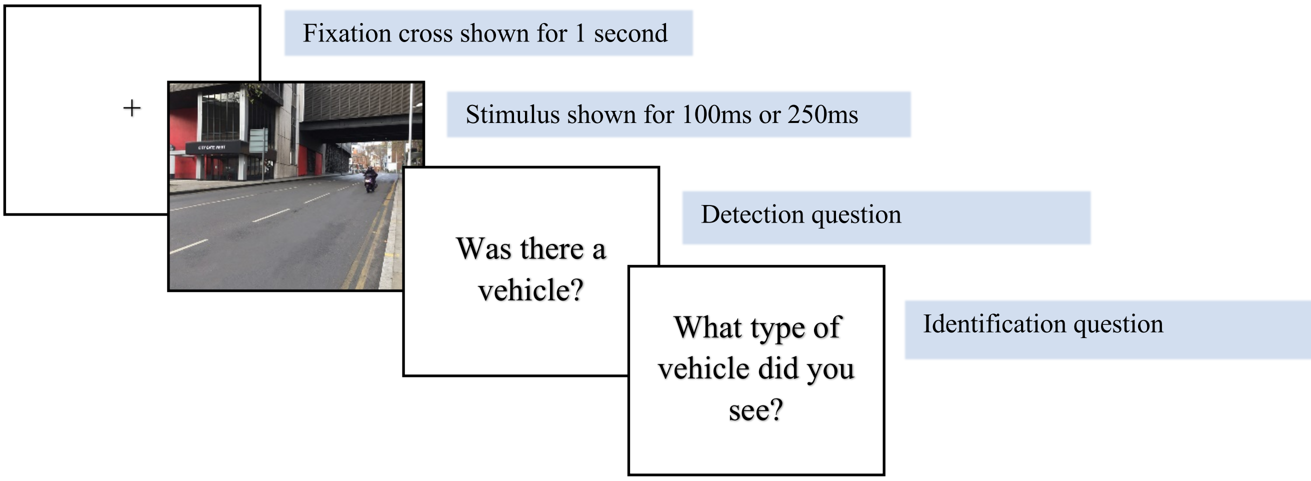

Figure 3.9: Summary of each trial beginning with the fixation cross, followed by the stimulus for either 100ms or 250ms, and then the two questions about the stimulus just presented.

Analysis

Data were analysed using the statistical programming language R (R Core Team, 2020). Detection accuracy and identification accuracy were analysed separately. Given that identification is dependent on accurate detection but not the reverse (Mack & Palmeri, 2010), analysis of identification accuracy was conducted only on trials where detection was correct. Response accuracy of each trial was analysed and was a binary variable (correct or incorrect). Trials with empty roads were not included in the analysis as the present study did not focus on false alarms in hazard detection. False alarm rates were low; 5.95% and 6.42% of empty road trials presented at 100ms and 250ms respectively were incorrect.

A linear mixed model with a binomial link function from the lme4 package was used to model detection and identification accuracy (Bates et al., 2015). The fixed effects were tested using likelihood ratio tests within the drop1 function from the lme4 package as described in

Chapter 2. The maximal random effects structure differed for each dependent variable and thus will be reported in their respective sections (see Barr et al., 2013). Age was means centred such that the intercept represents the mean accuracy for the average aged driver when all other variables are held constant at baseline.

3.3.3 Results

Detection accuracy

The maximal random effects structure for detection accuracy data accounted for random variance between stimuli and participants, as well as the within-participant variance for repeated variables. The full model with the specified random structure performed better than the traditional general linear model without random effects (see Table 3.6). The full model was:

Accuracy ~ Driver group + Vehicle type + Duration + Distance + Driver group:Vehicle type:Duration:Distance + Age + (1|Stimulus) + (Vehicle type|Participant) + (Duration|Participant) + (Distance|Participant)

Reference/Baseline categories of fixed effects: Driver group - car driver, Vehicle type – car, Distance – mid, Duration – short

Table 3.6: Model fit comparison (below significance threshold of .05)*

Source	**AIC**	**χ^2 difference**	***p***
Full model	2623.6	1542.6	< .001*
GLM model without random effects	4126.1		

The mean detection accuracy for each vehicle type and viewing condition according to driver group is shown in Figure 3.10.

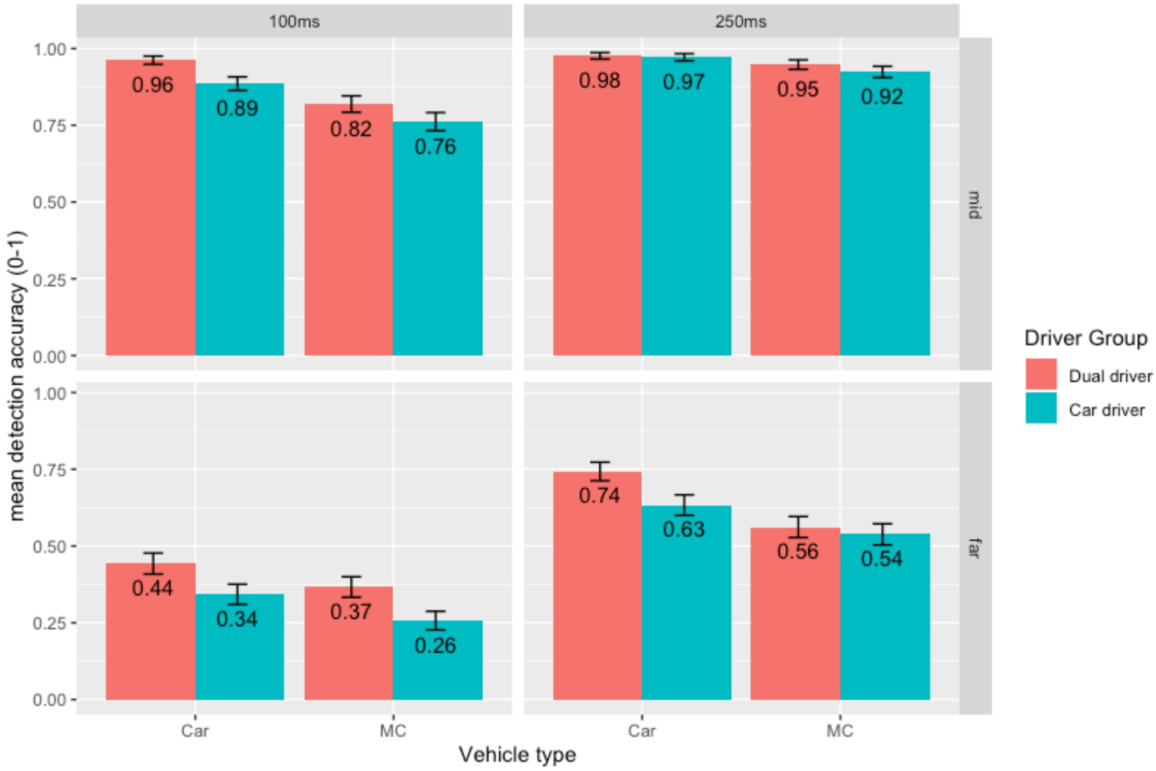

Figure 3.10: Mean detection accuracy from 0-1 with +1/-1 standard error bar

Applying the drop1 function on the main effects model, the likelihood ratio test found that driver group was a significant factor (χ^2 (1) = 5.00, p = .025). Dual drivers (M = 0.73, SD = 0.45) were on average more accurate than car drivers (M = 0.66, SD = 0.47; *OR* = 2.30 [95% CI: 1.10, 4.80]). Vehicle type was also a significant factor (χ^2 (1) = 10.11, p = .0014) with cars (M = 0.75, SD = 0.44) detected more accurately than motorcycles (M = 0.65, SD = 0.48; *OR* = 0.42 [95% CI: 0.24, 0.73]). Duration was also significant (χ^2 (1) = 29.31, p < .001). At short presentation durations (M = 0.61, SD = 0.49), detection accuracy was lower than at long presentation durations (M = 0.79, SD = 0.41; *OR* = 4.89 [95% CI: 2.73, 8.75]). Distance was a significant factor (χ^2 (1) = 89.13, p < .001), with far away vehicles (M = 0.49, SD = 0.50) detected less accurately than mid vehicles (M = 0.91, SD = 0.29; *OR* = 0.040 [95% CI: 0.022, 0.071]).

To determine whether dual drivers have a motorcycle-specific advantage at the detection stage, likelihood ratio tests were also conducted on interaction terms that include driver group and vehicle type factors. None of the interactions were significant (see Table 3.7).

Table 3.7: Summary of interaction analyses for detection accuracy

Interaction	Df	χ^2	*p*	*OR* [95% CI]
Driver group x Vehicle type	1	1.08	.30	0.78 [0.49, 1.25]
Vehicle type x Distance	1	0.94	.33	1.60 [0.56, 4.57]
Driver group x Vehicle type x Distance	1	0.62	.43	1.49 [0.55, 4.02]
Driver group x Vehicle type x Duration	1	0.64	.42	0.71 [0.31, 1.66]
Driver group x Vehicle type x Duration x Distance	1	2.84	.092	0.18 [0.023, 1.39]

Identification accuracy

Models were fitted for identification accuracy on trials where detection was accurate. The maximal random effects structure for identification accuracy data accounted for variance between stimuli and participants. The full model was:

Id Accuracy ~ Driver group + Vehicle type + Duration + Distance + Driver group:Vehicle type:Duration:Distance + Age + (1|Stimulus) + (1|Participant)

Reference/Baseline categories of fixed effects: Driver group - car driver, Vehicle type – car, Distance – mid, Duration – short

The full model performed better than a traditional general linear model without random effects (see Table 3.8).

Table 3.8: Model fit comparison (below significance threshold of .05)*

Source	AIC	χ^2 difference	*p*

Full model	977.38	188.8	< .001*
GLM model without random effects	1132.18		

The mean identification accuracy for each vehicle type and viewing condition according to driver group is shown in Figure 3.11.

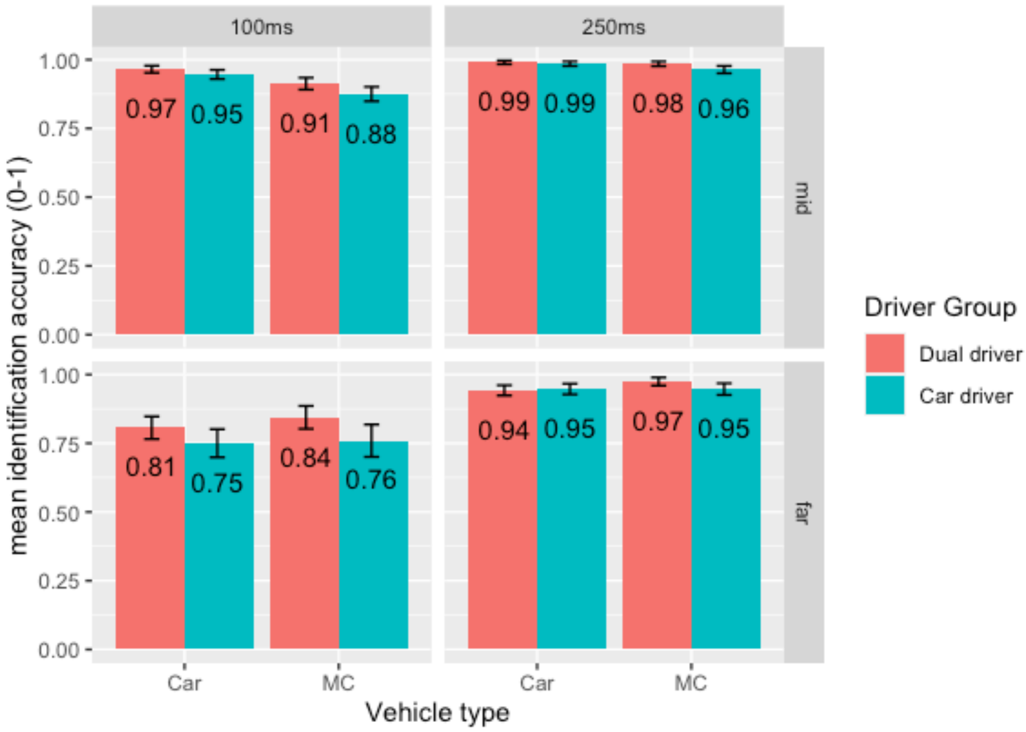

Figure 3.11: Mean identification accuracy from 0-1 with +1/-1 standard error bar

Applying the drop1 function on the main effects model, the likelihood ratio test found that driver group was a significant factor (χ^2 (1) = 5.08, p = .024). Dual drivers (M = 0.94, SD = 0.23) were on average more accurate than experienced car drivers (M = 0.93, SD = 0.26; *OR* = 2.14 [95% CI: 1.12, 4.11]). Duration was a significant factor (χ^2 (1) = 27.91, p < .001). At short presentation durations (M = 0.89, SD = 0.32), identification accuracy was lower than at long presentation durations (M = 0.97, SD = 0.17; *OR* = 5.76 [95% CI: 3.06, 10.8]). Distance was also significant (χ^2 (1) = 21.05, p < .001), with lower accuracy for far away vehicles (M = 0.90, SD = 0.31) than mid vehicles (M = 0.96, SD = 0.21; *OR* = 0.22 [95% CI: 0.12, 0.41]).

However, vehicle type was not a significant factor (χ^2 (1) = 1.52, *p* = .22; *OR* = 0.68 [95% CI: 0.37, 1.25]).

To determine whether dual drivers have an advantage for motorcycles at the recognition stage, likelihood ratio tests were also conducted on interaction terms that include driver group and vehicle type factors. None of the interactions were significant (see Table 3.9).

Table 3.9: Summary of interaction analyses for identification accuracy

Interaction	Df	χ^2	*p*	*OR* [95% CI]
Vehicle x Distance	1	2.62	.11	2.43 [0.78, 7.65]
Driver group x Vehicle type	1	0.85	.36	1.41 [0.67, 2.98]
Driver group x Vehicle type x Duration	1	0.21	.64	1.47 [0.28, 7.82]
Driver group x Vehicle type x Distance	1	0.29	.59	1.51 [0.33, 6.88]
Driver group x Vehicle type x Duration x Distance	1	0.068	.79	0.64 [0.021, 19.0]

3.3.4 Discussion

The current study presented cars and motorcycles at varying distances and durations in a T-junction task. The findings confirmed that unsurprisingly, vehicles presented further away and also for short durations were harder to detect and identify. The hypothesis for the main effect of driver group was also supported – dual drivers were more accurate at detecting and identifying vehicles than car drivers across both distances and durations. However, the findings did not support any of the interaction hypotheses involving driver group. There was no dual driver advantage for motorcycles specifically at detection and identification, and the hypothesised motorcycle-specific advantage for dual drivers was also not observed even under the most challenging conditions (i.e., far distance and short presentation durations). When vehicles were presented at far distances and for short durations, detection and identification accuracy deteriorated for both driver groups.

The aim of the current study was to investigate whether dual drivers have a motorcycle-specific processing advantage at detection and identification. The present study found no interactions between driver group and any other variables, and provides further evidence to Crundall et al.'s (2017) findings. The authors reported no significant driver group interactions in detection accuracy, and the present findings contribute to the literature by not only replicating but also extending this pattern of results to identification accuracy. Therefore, the evidence indicates that dual drivers do not have a lower processing threshold for detecting and identifying motorcycles. However, in the present study, there was a dual driver advantage for detection and identification of vehicles overall. This is in line with recent evidence that dual drivers have a visual processing advantage for cars and motorcycles in general (Crundall et al., 2017). Therefore, the findings suggest that dual drivers do not have a motorcycle-specific advantage at detection and identification but are better than car drivers at detecting and identifying both cars and motorcycles. The general advantage observed in dual drivers may be attributed to their motorcycle riding experience, as they become more vigilant not only of motorcyclists but also of other road users who might pull out in front of them. Therefore, dual drivers may not be better at spotting motorcyclists on the road per se, as they are more sensitive to all traffic.

Interestingly, the present study did not find an interaction between vehicle type and distance as in previous studies (Crundall et al., 2017; Crundall, Humphrey, & Clarke, 2008; Lee et al., 2015). Previous studies reported that while detection accuracy decreased with distance, detection accuracy of motorcycles declined significantly more so than cars. The inability to replicate the same pattern of results could be attributed to differences in statistical analyses. For instance, previous studies analysed the group average of accuracy percentage by driver group and distance whereas the current study analysed accuracy by trial. As mentioned in Chapter 2, analyses based on averaged scores can lead to inflation of Type I error (see

Baguley, 2012). By pooling together observations across participants, any sampling errors or variance from participants are ignored. Furthermore, the differences in variability between stimuli was ignored in previous studies. This means that previous results are potentially limited to the particular set of stimuli used in the experiments. When the current analysis was run again without accounting for variability among participants and stimuli, detection of motorcycles was indeed significantly worse than cars at the far distance[1].

The present study is novel because it measured detection and identification thresholds separately in relation to car and motorcycle hazards. Previous studies investigating motorcycle perception rarely differentiate the initial detection stage from subsequent processing stages that require further visual processing such as identification (e.g., Crundall et al., 2017; Crundall, Humphrey, & Clarke, 2008; Lee et al., 2015). By distinguishing detection from identification abilities, the present study was able to shed light on why motorcycles are particularly susceptible to LBFTS collisions. The present findings revealed that although motorcycles were less accurately detected than cars, there was no difference in identification accuracy upon detection. In other words, although motorcycles were harder to detect than cars, there was little difference in drivers' ability to identify the motorcycle or car once the vehicle has been detected. This suggests that perceptual failures around motorcycles may be due to difficulty in detecting the presence of motorcycles in the first instance, rather than difficulty in processing motorcycles. Once drivers have detected the presence of a motorcycle, they would have already identified that it is a motorcycle. In this particular instance, the current finding supports Grill-Spector and Kanwisher's (2005) observation that detection and identification of the object's basic-level category may be tightly linked. Therefore, the high rates of perceptual failures involving motorcycles may be due low detection rates rather than low post-detection identification rates.

[1] Without the inclusion of random effects for participant and stimulus in the model, the interaction between vehicle type and distance was significant (χ^2 (1) = 10.45, p = .0012). The decline in detection accuracy from mid to far distances was greater for motorcycles than cars (z = 3.16, p = .0015). At mid distance, odds ratio of detecting motorcycles to cars was 0.31 [95% CI: 0.21, 0.45] and at far distance, odds ratio of detecting motorcycles to cars was 0.62 [95% CI: 0.51, 0.76].

In conclusion, dual drivers did not exhibit an advantage specifically for motorcycles in terms of detection and identification abilities. They were affected by short time pressure and distance to a similar extent as car drivers. However, dual drivers are likely to have developed better visual processing skills that allows them to detect and identify vehicles more accurately in general.

3.4 General discussion

This chapter reported three experiments conducted to investigate whether dual drivers have a motorcycle-specific perceptual advantage. The experiments examined whether dual drivers have better hazard prediction abilities, quicker to detect motorcycles in a search array, and have lower detection and identification thresholds for motorcycles than car drivers. Across all three experiments, there was no motorcycle-specific perceptual advantage for dual drivers. Together with Crundall et al.'s (2017) findings, there is increasing evidence that dual drivers do not have a motorcycle-specific advantage in perceptual skills, but instead a general perceptual advantage. This was reflected in Experiment 3, which found that dual drivers were more accurate at detecting and identifying both cars and motorcycles than car drivers. This suggests that dual drivers' perceptual abilities may have developed to accommodate a wide range of hazards rather than specifically for motorcycles. The enhanced detection and identification abilities may stem from personal experiences of vulnerability on a motorcycle. To conclude, the notion that dual drivers are better than car drivers at spotting other motorcyclists on the road was not supported based on the three perceptual tasks used. Based on these findings, there is a need to re-assess the claim that dual drivers have superior motorcycle perceptual abilities compared to car drivers. An implication is that the lower collision rates with motorcycles observed in dual drivers compared to car drivers (Magazzù et al., 2006) may not be due to superior hazard perception and processing of motorcycles. However, the caveat is that we cannot be certain whether the dual drivers sampled in the three studies have lower collision rates than the car drivers sampled, as reported by Magazzù et al. (2006). Furthermore, the causal relationships between driver group membership and perceptual abilities, as well as between perceptual abilities and collision rates with motorcycles is uncertain. While dual drivers may have developed better hazard perception abilities due to their motorcycle riding experience, it is also possible that drivers with better hazard perception abilities are more inclined to learn to ride motorcycles. Similarly, there is yet to be evidence demonstrating that dual drivers' low collision rates with motorcycles is attributed to enhanced perceptual abilities for motorcycles rather than other factors such as cognitive schemas.

At the end of Chapter 1, the need for an effective perceptual training task targeted at improving drivers' ability to detect motorcycles while engaging drivers was discussed. The potential of a Pelmanism game-based training tool for motorcycle perception was identified (see Crundall et al., 2017) and the following chapters are dedicated to further developing, testing and evaluating this training tool. It was previously suggested that the training tool can be evaluated using perceptual tasks reported in this chapter. The initial rationale was to first assess dual drivers' performance on these perceptual tasks before using them to measure the impact of the motorcycle Pelmanism training task. If trained participants performed better on the same perceptual tasks that were sensitive to drivers who are knowledgeable about motorcycles, this would ensure that the participants developed skills associated with the experiential differences in motorcycle perception. Although the tasks used in the current chapter did not find any dual driver advantage specific to motorcycles, demonstrating a dual driver advantage, while sufficient, was not necessary to conclude that the perceptual tasks are appropriate for measuring motorcycle perceptual skills. Furthermore, previous research has demonstrated differences between driver groups that are expected to have different levels of perceptual skills, such as novice drivers and experienced drivers, in hazard prediction, visual search and detection abilities (e.g., Crundall et al., 2012; Sanocki et al., 2015; Ventsislavova & Crundall, 2018). Therefore, the perceptual tasks can still be used to evaluate any improvements in motorcycle perception skills before and after the proposed perceptual training task. In the next chapter, a second literature review is presented, which discusses existing training interventions and serious games.

Chapter 4 – Introduction to game-based training

Chapter overview

This literature-review chapter provides an overview of research related to existing road safety interventions and serious games designed for learning or skill acquisition. The first half of the chapter reviews the literature on existing road safety interventions and whether they are effective in mitigating perceptual failures. Existing road safety interventions include awareness campaigns and training interventions that are designed to improve drivers' hazard perception ability. However, these interventions do not address the propensity of perceptual failures involving motorcycles and they also lack appeal for drivers to take part in the training intervention. A game-based training intervention that has demonstrated potential to improve the abilities of drivers to detect motorcycles is subsequently discussed. The second half of the chapter thus reviews the literature on serious games, which are games that are designed for learning and skill acquisition rather than entertainment purposes, in order to provide some background information on the game-based training method targeted at improving motorcycle detection.

4.1 Current road safety interventions

4.1.1 Awareness campaigns

In the recent years, the British government has focused on awareness campaigns that encourage driver empathy towards motorcyclists while reminding them to be vigilant for motorcyclists at junctions. As part of the THINK! Road safety campaigns, the government has introduced several THINK! Bike campaigns such as 'Think Bike, Think Biker!' and 'Didn't see' (Department for Transport, 2020a). These THINK! Bike campaigns addressed various driver safety behaviours such as checking blind spots for motorcyclists and taking longer to check for motorcyclists at junctions. The campaigns also attempted to create driver empathy by humanising motorcyclists in adverts, rather than simply presenting statistics. Motorcycle collisions resulting from perceptual errors were specifically addressed in the 2014 THINK! Bike campaign. For example, the difficulties of detecting a motorcycle were highlighted on a poster (see Figure 4.1).

Figure 4.1: Poster used in the THINK! Bike campaign in 2014 (adapted from Department for Transport, 2020b). Permission to reproduce this figure has been granted by the [Open Government Licence](#) for [Crown Copyright](#)

One of the campaign objectives was to remind drivers that the main cause of motorcycle collisions is drivers' failure to look properly, especially at T-junctions. The intention of the poster was to demonstrate the difficulty of detecting motorcycles. However, the poster received backlash for adjusting the contrast to artificially reduce the visibility of the motorcycle, which may have limited its effectiveness in changing driver attitudes towards motorcycles (Visordown News, 2014). The effectiveness of motorcycle awareness campaigns, particularly the 2014 campaign, is discussed below.

Limitations of awareness campaigns in addressing motorcycle perceptual failures

An unintended consequence of the poster in the 2014 campaign was that drivers interpreted the poster as an excuse for failing to detect motorcycles and reinforced their view that motorcyclists should be responsible for their own visibility (Visordown News, 2014). By over-exaggerating the difficulty of detecting motorcycles, drivers may have felt absolved from their role in motorcycle collisions. A survey conducted after the 2014 THINK! Bike campaign found little to no change in drivers' attributions of motorcycle collision causes (TNS BMRB, 2014). Apart from cases where participants in the survey did not attribute responsibility, car drivers were most likely to attribute motorcycle collisions to speeding motorcyclists. This indicated that drivers still considered motorcyclists to be reckless and responsible for their involvement in collisions. Only when participants were prompted using a list of possible causes did they attribute collisions to drivers' failure to yield appropriately at T-junctions (TNS BMRB, 2014). The continued attribution of motorcycle collisions to the fault of motorcyclists is problematic and increased effort is required to address the contributory role of car drivers in motorcycle collisions.

The limited effectiveness of motorcycle awareness campaigns in changing drivers' attitudes towards motorcyclists was also supported by findings from focus group research conducted by Musselwhite et al. (2012). The study found that although the public thought that the THINK! Bike campaigns were memorable, there was still a significant lack of consideration and mutual awareness between car drivers and motorcyclists (Musselwhite et al., 2012). Notably, car drivers were still adamant that motorcyclists should be responsible for their involvement in collisions, and that fluorescent clothing or daytime running lights (DRL) on motorcycles would be sufficient to prevent collisions with motorcycles. This demonstrates that car drivers still held misconceptions about motorcycle collisions and were still unwilling to acknowledge that car drivers are responsible for the majority of collisions involving motorcycles. As mentioned in the first chapter, collisions with motorcycles often occur because a driver failed to give way appropriately and in the majority of the cases, this was due to a failure to perceive the motorcycle in time (Clarke et al., 2004, 2007; Department for Transport, 2015a). Furthermore, collisions involving motorcycles still occur even in high visibility conditions and also when the motorcyclist was wearing fluorescent clothing (ACEM, 2009; Clarke et al., 2007; see also Langham et al., 2002).

While effective awareness campaigns may have a role to play in changing drivers' attitudes by drawing attention to the risks that vulnerable road users such as motorcyclists face, raising awareness is not sufficient to improve drivers' ability to detect motorcycles on the road. A more direct approach to reducing perceptual failures by drivers is through driver training interventions. Training interventions can provide drivers with opportunities to learn and practice skills that enable safe driving and may offer a better solution to perceptual failures leading to LBFTS collisions than awareness campaigns.

4.1.2 Hazard perception training

Hazard perception is a crucial driving skill that has been associated with collision rates. For this reason, researchers have investigated various training methods to improve drivers' hazard perception skills. In the section below, I will describe the different strategies used in hazard perception training interventions. The training methods include having drivers practise predicting what happens next, listen to commentary provided by expert drivers during hazard perception, or to provide their own commentary of different hazard scenarios.

Hazard prediction training

Previous research has shown that hazard perception ability improves with driving experience and this is presumably because experience creates a more cognisant mental model of the driving environment that facilitates hazard perception (see Horswill & McKenna, 2004). Experienced car drivers and drivers without history of collision involvement have been shown to be better at interpreting environmental cues to help predict what will happen next as compared to novice car drivers and drivers who had been involved in collisions previously (e.g., Crundall, 2016; Jackson et al., 2009; Ventsislavova et al., 2016). Consequently, hazard prediction tests are not only used to measure hazard perception ability but also as a training method to improve drivers' hazard perception ability (e.g., Wetton et al., 2013). The objective of hazard prediction training is to train drivers to generate all potential outcomes based on the available environmental cues, in order that they can develop better insights to the traffic situation. In this training method, drivers are shown several traffic clips that are cut immediately before the hazard occurs and are instructed to generate all possible outcomes based on the visible environmental cues. They are then given feedback and a walkthrough of the relevant cues in that particular situation. There is some evidence that hazard prediction training may improve hazard perception. In Wetton et al.'s study (2013), participants who completed a hazard prediction training task subsequently detected hazards quicker in a hazard perception test than participants who completed a control training task where they watched a driver safety video.

Similar to hazard prediction training, Meir et al. (2014) proposed a hazard perception training known as the Act and Anticipate Hazard Perception Training (AAHPT), in which novice drivers were exposed to a range of traffic hazards (see also Fisher et al., 2006). Participants were randomly allocated to one of four groups – an active training group, instructional training group, hybrid of active and instructional training group, or control group. In the active group, participants were shown video clips and had to press a button each time they detected a hazard, whereas in the instructional training group participants took part in a passive tutorial on hazard perception, and the control group watched a road safety tutorial. A week later, participants' eye movements and reaction time on a hazard perception test were recorded. The study found some benefit for the hybrid training group, but this was limited only to certain categories of hazards such as pedestrian hazards. Although a conclusion may not be drawn from the limited studies conducted, the lack of research evaluating the effectiveness of hazard prediction training specifically is stark, as studies often combine hazard prediction exercises with other forms of training (Chapman et al., 2002; Horswill, Taylor et al., 2013).

Listening to expert commentary

Another type of hazard perception training is to present a video of a developing hazard with commentary produced by a driving instructor or someone of similar level of expertise. The commentary provides insights to the relevant environmental cues and teaches drivers how to recognise and anticipate potential hazards as the hazard develops in the video (e.g., Wallis & Horswill, 2007). The objective was to help direct drivers' attention to the appropriate locations during driving so as to improve their ability to perceive hazards. Novice drivers who received training by listening to expert commentary were found to perform better in a subsequent hazard perception test measured using reaction time compared to novice drivers who watched the video without expert commentary (Wallis & Horswill, 2007). Training benefits were also found for older drivers above the age of 65 (Horswill et al., 2010) and even experienced car drivers with an average of 30 years of driving experience (Horswill, Taylor et al., 2013). Using a hazard prediction test to measure hazard perception ability without the influence of appraisal (see Crundall, 2016), Castro et al. (2016) provided further support that improvements after expert commentary training could be attributed to actual improvements in hazard perception ability rather than changes in response criterion. In Castro et al.'s (2016) study, trained drivers who received the expert commentary training produced more accurate and detailed predictions of the developing hazard than untrained drivers.

Despite the body of cumulative evidence described in the preceding paragraph, the effect of expert commentary training is not to be taken for granted. As Young et al. (2017) pointed out, the training effect may simply be due to a difference in engagement levels between the control and trained conditions. In previous studies, participants in the control condition were instructed to watch the video passively and they may not have been motivated to engage with the video properly. When the level of engagement in the control training condition was increased by having participants press a button whenever they saw a hazard, there was no benefit of expert commentary training (Young et al., 2017). Therefore, it is possible that listening to expert commentary simply engages the driver with the traffic scene and developing hazards during training. Commentary training is not limited to listening to expert commentary, because drivers may be inclined to also produce their own commentary after having learnt how commentary should be produced to facilitate hazard perception. The following section describes the use of self-generated commentary as a method of hazard perception training that is often used in combination with expert commentary.

Self-generated commentary

A similar strategy to the expert commentary training is to ask drivers to generate their own commentary while being exposed to different traffic scenarios. Similar to expert commentary training, the underlying objective is to direct drivers' attention to hazard precursors and help them better anticipate hazards. However, self-generated commentary training emphasizes the role of driver engagement during training as drivers have to actively process the environmental cues and learn how different hazards develop through their own experience. For example, drivers are encouraged to practise identifying hazards aloud and predicting what might happen next. This is based on the idea that experiential training is often more effective than instructional training (Meir et al., 2014). Isler et al. (2009) tested the effectiveness of self-generated commentary training on novice drivers by measuring their ability to identify hazards while performing a secondary tracking task to simulate driving. Novice drivers who took part in the commentary exercise subsequently identified more hazards than those who did not take part. Other studies using a hazard perception test and driving simulator measures to evaluate the effect of training also reported a positive training effect (Crundall, et al., 2010; Cantwell et al., 2013). However, these studies did not specify whether participants are required to generate their own commentary during the post-training evaluation task, in addition to the training task. Evidence suggests that the training effects of self-generated commentary disappears when participants are also required to generate commentary during the post-training evaluation task (Young et al., 2014; Young et al., 2017). This suggests that concurrent commentary may interfere with hazard perception.

Surprisingly, combining expert commentary and self-generated commentary may inadvertently limit their effectiveness as a training method. Young et al. (2017) also investigated whether the detrimental effect of concurrent commentary production on hazard perception would persist with and without prior exposure to an example expert commentary. They found that concurrent commentary negatively affected hazard perception only after listening to expert commentary, even with commentary practice. The authors proposed that providing trainees with a reference may have increased the cognitive workload of having to produce commentary of a similar standard to the reference. For instance, having to express their reactions to hazard precursors and rationalise each action is likely to place additional demands on the hazard perception task. In contrast, spontaneous commentary is not restricted by expectations derived from listening to experts' commentary. The negative effect of combining expert and self-generated commentary training is supported by another study, which found that adding self-generated commentary training to expert commentary training had little impact on subsequent improvements in hazard perception training (Wetton et al., 2013).

So far, the three approaches to hazard perception training described are hazard prediction training, expert and self-generated commentary training. A comparison of the three hazard perception training methods was conducted by Wetton et al. (2013). Participants were randomly allocated to one of the four training conditions – hazard prediction training, expert commentary training, a hybrid of expert and self-generated commentary training, and a full training consisting of hazard prediction training and the hybrid commentary. The study found that while all training conditions improved hazard perception ability, the hazard prediction training was the least effective and its training effects diminished the most after a period of delay. In contrast, the full training resulted in the largest improvements in hazard perception ability. Although these approaches to hazard perception training have reported some positive training effects, they may have limited effectiveness in reducing perceptual failures involving motorcycles for the reasons outlined in the next section.

Limitations of hazard perception training in addressing motorcycle perceptual failures

While the training methods described above are promising for improving overall hazard perception ability, they may be less effective for improving drivers' ability to perceive motorcycles on the road. Current hazard perception training methods are aimed at improving hazard perception of general on-road hazards and a common limitation is that they lack emphasis on motorcycles as a source of hazard. Given that motorcyclists are disproportionately affected by perceptual failures by other drivers (ACEM, 2009; Clarke et al., 2004, 2007; Department for Transport, 2015a), there is a need for a motorcycle-specific training intervention aimed at improving drivers' perceptual skills for motorcycles. The rationale for a motorcycle-specific perceptual training intervention is supported by evidence demonstrating that target-specific perceptual training is effective in increasing hit rates and reducing false alarms in target detection (Guznov et al., 2017). Through repeated practice and exposure to the target stimuli, perceptual training can induce long-lasting changes to visual neurons and mental representations of the target such that subsequent perception is improved (Karni & Bertini, 1997; Sagi, 2011; Sasaki et al., 2010). Therefore, there is scope in the literature to develop a perceptual training tool that improves drivers' perceptual skills for detecting motorcycles.

A second key limitation of existing methods of hazard perception training is the lack of incentive for drivers to take part and engage in the training. Drivers may be reluctant to engage with hazard perception training because they believe their hazard perception skills to be good enough already. There is a plethora of research documenting the 'better than average' effect in drivers, as drivers were often found to rate their hazard perception skills to be better than the average UK driver as well as others with the same driving experience as themselves (Horswill et al., 2004; Horswill, Sullivan et al., 2013). Although experts are better at monitoring their own performance than non-experts, the 'better than average' effect was found even when expert drivers were told to rate amongst themselves, suggesting that self-enhancement does not diminish with experience or training (Waylen et al., 2004). If drivers are not willing to take part in training interventions voluntarily, the rate of training uptake will be low. This poses a major problem for training interventions because the real-world impact of the training is subsequently limited, and any training benefits found in studies would unlikely be generalised beyond research settings. Therefore, the lack of appeal poses a real problem for training interventions and needs to be addressed in the proposed motorcycle-specific perceptual training tool. Based on the two key limitations highlighted, there is scope to explore a different form of training method that addresses both the lack of motorcycle-specific training intervention and appeal.

4.2 Serious games as perceptual training

The need for an engaging training task targeted at improving drivers' ability to perceive motorcycles can be resolved by delivering a game-based motorcycle perceptual training task. Game-based training is also known as serious games, which are defined as games designed primarily to improve certain skills or knowledge rather than for entertainment purposes (Loh et al., 2015). Crucially, the game itself is the training and the game is fully integrated with the training content to facilitate learning. This is in contrast to gamification, which uses elements of games to improve user experience in non-game contexts (see Deterding et al., 2011). It is important to clarify a common misconception that unlike serious games, gamification does not refer to games and gamification of an application does not transform it into a game (Loh et al., 2015). This thesis focuses on serious games rather than gamification. The ultimate purpose of serious games is motivated learning, where learning outcomes are achieved through self-motivated and repetitive play (Garris et al., 2002). This is so that players are motivated to take part in the training voluntarily and willing to keep doing so instead of a one-off training session. Therefore, there is an advantage for delivering perceptual training as a serious game, because both training objectives and motivational outcomes can be achieved through proper game design.

The design of serious games consists of a learning aspect as well as a motivational aspect and they are achieved through game mechanics and game elements respectively. Both aspects contribute to the experience of gameplay for players. Game mechanics refer to the rules, procedures, actions that players can perform as well as content that make up the core of the game, whereas game elements refer to features that engage and improve the experience of gameplay for the player. Game elements are therefore decorative features that enhance the game, but they cannot sustain or run the game on their own. The game mechanics need to be driven by intended learning outcomes or training objectives, otherwise it would simply be a game for entertainment (Charsky, 2010). Similarly, the use of game elements needs to be informed by motivational theories that complement the training objectives, in order to avoid distracting users from the intended training objectives. Having too many decorative game elements could backfire and overload the player's cognitive capacity to learn from the content, thus failing to achieve the intended training objective (Gunter et al., 2006; Kinzer et al., 2015). Therefore, an understanding of the desired motivational outcomes and how to achieve them in serious game design is needed.

4.2.1 Desired motivational outcomes

As mentioned previously, the gameplay experience can be enhanced through game elements to achieve repetitive and self-motivated play (Garris et al., 2002). Game elements facilitate learning by keeping players engaged and motivated with the learning content. A primary characteristic of gameplay is that players become absorbed with the game and do not turn the game off willingly (Garris et al., 2002). This is often referred to as flow, which is a subjective and optimal state of enjoyment, in which players are operating at full capacity and forget about their surroundings. The concept of flow was developed by Csikszentmihalyi (1990) and is often cited as a desired outcome of serious games (Grund, 2015). The state of flow is characterised by an intense concentration on the present activity, loss of awareness of surroundings, feelings of time distortion where time seems to pass faster than normal, merging of actions and awareness where actions are performed almost automatically, and an intrinsically rewarding experience (Nakamura & Csikszentmihalyi, 2002). Experiencing flow therefore enhances players' experience during gameplay, which in turn motivates repetitive gameplay because the game becomes enjoyable and rewarding on its own. Without a positive gameplay experience, players are less likely to be self-motivated to return to the game and may only play it for external factors such as money or avoidance of punishment (Garris et al., 2002). Therefore, it is important for the proposed motorcycle perceptual training tool to be able to induce a flow state in players so that they may keep engaging with the learning content.

The need for challenges

There are certain conditions that need to be met before players enter a state of flow. First, the activity must present challenges that are pitched at the appropriate level to the players' capabilities (Nakamura & Csikszentmihalyi, 2002). The relationship between challenge and enjoyment was suggested to take the form of a curvilinear inverted U-shape (Abuhamdeh & Csikszentmihalyi, 2012). Initially, increases in task difficulty enhances enjoyment up until a certain point, after which further increases in difficulty results in frustration and decrease in enjoyment. The importance of sufficiently challenging game elements is also echoed by the self-determination theory (Ryan & Deci, 2000). Applying this theory to gameplay, players feel intrinsically motivated to play a game when their need for competence is fulfilled through appropriately difficult challenges and opportunities to demonstrate their skills (Ryan et al., 2006). Besides competence, Ryan and Deci (2000) also suggested that perceived autonomy is required to facilitate intrinsic motivation. Perceived autonomy refers to the sense of control over one's behaviours and in a game context, this means that players need to experience freedom in choosing and performing actions within the game. In the absence of perceived autonomy, players may attribute their motivation and level of effort to external regulators rather than self-driven motivation, thus undermining intrinsic motivation (van Roy & Zaman, 2017). Consequently, van Roy and Zaman (2017) proposed that serious games should provide a range of challenges that players can choose from, otherwise restricting players' autonomy may subsequently diminish their intrinsic motivation.

Based on the first condition of flow and the self-determination theory, a potential game element would be having different levels of difficulty. As the optimal level of challenge differs between individuals and is based on the individual's level of skills, the levels of difficulty should also match the players' skill level to be effective in creating an optimal gameplay experience. To achieve this, some researchers suggest that the levels of difficulty could increase gradually throughout the game, starting off easily and increasing as players develop mastery of the game (Bostan, & Öğüt, 2009; Desurvire et al., 2004; Sweetser & Wyeth, 2005). This would maintain a certain level of challenge throughout the game. However, as the optimal level of challenge is likely to differ between individuals, difficulty levels may be more effective if they could be adjusted based on the player. Although recent developments in algorithm have allowed the level of difficulty in the game to adapt to players' performance, empirical evidence suggests that there is actually little advantage of an automated adaptive design over pre-defined increments of difficulty levels (e.g., Sampayo-Vargas et al., 2013; Smeddinck et al., 2016). There was little difference in players' engagement levels between an automatic adaptive version and a version with incremental difficulty. This could be because in automatic adaptive designs, players likely realised that they could never 'beat the game', which undermines their perceived autonomy and intrinsic motivation to keep playing.

Another potential game element that introduces challenges could be competition, with opponents of similarly matched skills. A study paired chess players with opponents of varying skill levels and at the end of the game, players were asked to rate their game experience with and without reference to the game outcome (Abuhamdeh & Csikszentmihalyi, 2012). Reference to the game outcome was controlled for because players are likely to have a more positive game experience if the outcome was in their favour than if it was not. The study found that players' enjoyment of the game was higher when paired with opponents of higher skill level than with opponents of lower skill level, even without reference to the game outcome. Enjoyment was also increased when players outperformed their opponents by a small margin rather than a wide margin, further indicating the importance of a suitably challenged game (Abuhamdeh & Csikszentmihalyi, 2012). This is in line with another study examining the effects of competition with higher, lower or equally skilled opponents (Santhanam et al., 2016). It was found that only players who faced an equally skilled competitor reported enhanced engagement with the game.

However, game designers must be cautious when using competition as a game element. This is because competition can have undesirable effects depending on the nature and goal of the competition. Competitions where the goal is to win and beat other opponents can be detrimental because it encourages extrinsic motivation, which can diminish motivation in the long run when reinforcements are no longer present (Fülöp, 2009). In addition, excessively competitive game elements could negatively impact learning by increasing anxiety from social comparison and can further discourage low-performing learners (Cheong et al., 2014; Cheng et al., 2009). Anxiety resulting from the fear of being evaluated negatively is known as evaluation anxiety or fear of negative evaluation when referring to social evaluations (Watson & Friend, 1969; Zeidner & Matthews, 2005). When competition overly focuses on social comparisons, the increased opportunity for evaluative situations in turn increases the anxiety felt by players during the competitive game. For example, students' self-reported levels of anxiety increased when exposed to evaluative situations such as working in groups or being called on to answer questions during lectures (England et al., 2017; see also Cooper & Brownell, 2020). Game elements that introduce excessive social comparison may unintentionally become stumbling blocks to learning. Therefore, competitions must be constructive and encourage intrinsic motivation. According to Fülöp (2009), constructive competitions must focus on the player's self-improvement and promote goals aimed at learning or gaining mastery of a task. In constructive competitions, the role of opponents is limited and only functions to facilitate the individual's self-improvement process.

The need for clear goals with goal-related feedback

The second condition for flow state is that challenges must be framed in terms of clear attainable goals and continuous feedback must be provided in relation to the goals (Nakamura & Csikszentmihalyi, 2002). This is in line with the goal-setting theory, which suggests that challenging and specific goals are more effective at improving motivation and performance than do-your-best goals (Locke & Latham, 2002). This is because goals that specifically outline the end goal or outcome act as tangible targets to work towards. The goal-setting theory also emphasizes the importance of providing feedback on progress in relation to the goal so that learners can adjust the level of effort or strategy used in order to reach the goal. Therefore, players need to be able to see their progress in relation to the goal throughout the game.

However, the effectiveness of goals is also contingent on players' commitment to the goal. According to the goal-setting theory, goal commitment mediates the impact of goals on motivation (Locke & Latham, 2002). If learners are not convinced by the goal set, they are less likely to be motivated to attain it. While making public commitments is one method of goal commitment, this could elicit negative performance anxiety and pressure. Instead, researchers like Latham (2004) and Lunenburg (2011) suggest that learners should be involved in the goal-setting process so that goals are perceived as personally important. Furthermore, by inviting learners to set their own goals, it ensures that the goal is not too difficult. If learners are given goals that are too difficult for their capabilities, it could lower their self-efficacy, which refers to the individual's confidence in their ability to attain a goal. In contrast, promoting self-efficacy can help increase the amount of effort players put in and persistence with attaining the task. As a result, many serious games attempt to enhance players' self-efficacy by rewarding successful attainment of goals in the form of badges or points (Mutter & Kundisch, 2014).

Badges and points serve as reminders of players' progress throughout the game and could further enhance their satisfaction of achieving their goal. This is in line with Bandura's (1982) notion that the most effective way to enhance self-efficacy is through mastery experiences or successful performance attainments, which can be easily represented in games as badges or points. Therefore, game elements that reward players, or game incentives, are often used in game designs to facilitate goal setting and reward goal attainment (Richter et al., 2015). For example, badges can be collected by players for successful completion of distinct goals and represent the players' past achievements in the game. Furthermore, they serve as a form of feedback by marking significant milestones reached in the game. The fact that badges are often aesthetically pleasing also contributes to players' desire to earn these collectables (Zichermann & Cunningham, 2011). Apart from being achievement markers, badges also signify a players' status or reputation in the game – by displaying his/her badges, others can infer information about the player's experience and skill level in the game. Points are similar to badges in that they also help facilitate goal setting. When points are awarded for accomplishing goals, players receive a direct measure of progression or even performance, which promotes feelings of self-efficacy. An advantage of points over badges is that they can be applied to many contexts, from scoring systems, progress indicators to virtual currencies (Zichermann & Cunningham, 2011). In scoring systems, players earn points based on their performance and can be combined with competition, resulting in high scores and/or leader boards. The use of points as progress indicators are often known as experience points in games, because it reflects the players' level of experience in the game. Generally, experience points do not have a limit or get depleted such that the longer a player has been playing the game, the more experience points they would have accumulated. Due to their ability to facilitate goal setting and enhance self-efficacy, game incentives are a powerful game element to motivate players and induce a state of flow.

Unfortunately, some researchers warn against potential negative effects of rewarding players in serious games in the long run. Although the goal-setting theory and self-efficacy theory suggest that game incentives can improve self-efficacy and encourage goal attainment, the self-determination theory argues that rewards could undermine players' intrinsic motivation. A longitudinal study focusing on the effects of collecting badges and coins on academic performance found that students' satisfaction and motivation to engage with the course decreased over time (Hanus & Fox, 2015). In comparison, students who did not collect badges and coins as part of the course became more motivated. Similarly, an investigation into the use of points and badges in Stack Overflow, which is a question-and-answer site for programmers, found that the incentives had unintended consequences for users' engagement and contribution to the site (Mamykina et al., 2011). For instance, points were rewarded for answering a question, with more points rewarded for quicker responses, and the number of points earned can grant users additional privileges, such as access to more tools. As a result, some users tended to provide shorter answers at the expense of details so that they can answer more questions in a shorter amount of time and do so before other users respond, thus earning more points. The study also found a dramatic reduction in participation on the site when users reached 10000 points, after which they no longer receive additional benefits. This demonstrates that incentives can undermine motivation over time, particularly when the rewards are no longer appealing.

However, these studies examined the use of game incentives within gamification, which differs from serious games. As mentioned previously, gamification is the use of game elements in non-game contexts. It is possible that intrinsic motivation is undermined not by the presence of rewards per se, but how rewards are used. In gamification, rewards are often offered in exchange for a certain behaviour, such as answering questions posted by others, which may be interpreted by users as part of a transactional deal. The use of rewards to control users' behaviours is likely to result in lowered perceived autonomy, which plays an essential role in intrinsic motivation (Ryan & Deci, 2000). In contrast, rewards that provide feedback or evidence of past successes may be effective in enhancing motivation. Furthermore, if players have the autonomy to set their own goals, and by extension the reward they would receive, rewards may not undermine intrinsic motivation during gameplay.

So far, a review of the conditions for achieving flow state provides some background information on what the proposed game-based perceptual training tool should entail. Motivational theories underlying game elements that are commonly used to introduce challenge and incentivise players such as difficulty levels, competition, badges and points, were also reviewed. In an attempt to develop an engaging training aimed at reducing perceptual failures with motorcycles, Crundall et al. (2017) proposed the use of Pelmanism to deliver motorcycle-specific perceptual training. The following section introduces the game of Pelmanism and reviews the theoretical bases for the underlying game mechanic and the potential for using motivating game elements in the Pelmanism-based perceptual training tool.

4.3 Pelmanism as a method of perceptual training

The objective of developing a motorcycle-specific perceptual training tool is to improve drivers' ability to detect motorcycles. As detection and identification abilities are enhanced for experts within their domain of expertise compared to non-experts (e.g., Hershler & Hochstein, 2009), it is possible to train drivers to improve on their perceptual skills, but how should training be conducted? There is evidence suggesting that training conducted at a subordinate-level but not at a basic-level improves perceptual skills (e.g., Scott et al., 2006). Subordinate-level training refers to training conducted using within-category stimuli whereas basic-level training refers to training conducted using between-category stimuli (more details later). A Pelmanism task that requires participants to match images of motorcycles on a subordinate level should therefore lead to improved perception of motorcycles and therefore improved detection of motorcycles (see Crundall et al., 2017). The evidence for perceptual expertise and subordinate-level training will be reviewed in further detail in the next section, after providing some background information on Pelmanism.

Pelmanism is a memory game that involves finding matching card pairs and is sometimes also known as Concentration or Pexeso. At the beginning of the game, all the cards are face down and players have to find matching card pairs by turning over only two cards at a time. If the two cards are a match, they are to remain faceup, otherwise they are returned facedown. There are many variants to the rules of Pelmanism, depending on the number of players and how players are scored. For example, Pelmanism can be played with a single player or two players taking turns to turn over two cards. The game can also be played to find as many pairs as possible within a fixed amount of time or to find all the pairs within the shortest possible time. In some variations of Pelmanism, players have to find matching cards of the exact same image (Figure 4.2) and in other variations, players may have to match words and pictures or cards that represent the same concept (Figure 4.3). If all the cards used in the Pelmanism game belong to the same basic-level category, for example 'dogs', then players have to make within-category discriminations or subordinate-level categorisations in order to make a successful match.

this image has been removed by the author for copyright reasons

Figure 4.2: Example of Pelmanism using card pairs that match exactly

this image has been removed by the author for copyright reasons

Figure 4.3: Example of variation in matching card pairs

4.3.1 Theoretical bases for the Pelmanism as perceptual training

The training objective of the proposed motorcycle perceptual training tool is to improve drivers' ability to detect motorcycles. Therefore, the game mechanic of Pelmanism must be aimed at helping drivers to develop perceptual expertise in motorcycles. Research has shown that people with expertise in a particular area show improved abilities to detect, identify, and categorise objects within their area of expertise, compared with non-experts. For instance, Hershler and Hochstein (2009) reported that experts were quicker and more accurate at detecting objects within their domain of expertise than non-experts in an array of distractor objects. This advantage was also replicated using objects embedded in natural scenes (Reeder et al., 2016). In another study, Bukach et al. (2010) demonstrated car experts were better at identifying whether a car was the same or different to a previously presented car, compared to non-experts. Interestingly, experts can also categorise objects within their expertise at a subordinate level as quickly as at basic level (Tanaka & Taylor, 1991). For example, dog experts can categorise dogs as quickly as they can categorise Labradors and are also more likely to use subordinate labels (Labrador) during basic-level categorisation than non-experts (see Figures 4.4 and 4.5 for pictorial example).

this image has been removed by the author for copyright reasons

Figure 4.4: Basic level categorisation – which is the dog? Basic-level categorisation should be easy for experts and non-experts, but dog experts are more likely to refer to the dog as a Labrador.

this image has been removed by the author for copyright reasons

this image has been removed by the author for copyright reasons

Figure 4.5: Subordinate level categorisation – which is the Labrador? Dog experts are equally quick at discriminating between the Labrador (left) from Golden retriever, as basic-level categorisation (Figure 4.4).

As experts have access to domain-specific knowledge, researchers have suggested that experts are able to extract visual information relevant to the task easily (Harel, 2016; Harel et al., 2010). According to Ullman (2007), this domain-specific knowledge involves a set of features that facilitate between-category and within-category discriminations. Experts can therefore recognise features that are representative of the object category and also take into account variability in how features appear between objects within the same category. Although real world perceptual expertise is usually developed over years through experience, the development of perceptual expertise could potentially be accelerated through repeated and deliberate exposure to relevant perceptual information. A motorcycle perceptual training task aimed at developing perceptual expertise can therefore improve drivers' ability to detect motorcycles.

There is a body of evidence indicating that training conducted at a basic level is not effective in improving participants' perceptual skills in terms of detection, but training conducted at a subordinate level is. Archambault et al. (1999) compared the effectiveness of training conducted at a subordinate level and basic level using a change detection task while controlling for location of the change. The study found that participants detected changes to objects trained using specific labels more quickly than objects trained using basic category labels (e.g., Mary's mug vs mug respectively). This suggests that the subordinate-level training task enhanced how the trained object was encoded and stored in memory, thereby allowing observers to compare the encoded image with visual input from the changed scene (see Hollingworth, 2003). This was also evident in applied contexts, such as drone operation, where operators who were trained to identify specific military targets at a subordinate level demonstrated improved search accuracy during a simulated flight (Guznov et al., 2017). Similarly, training conducted at a subordinate level is also more effective than training conducted at a basic level in improving participants' ability to identify trained objects. A study conducted by Scott et al. (2006) and another study conducted by Tanaka et al. (2005) trained a group of participants to identify bird species at a subordinate level (e.g., screech owl) and another group to identify bird families at a basic level (e.g., owl). In both studies, the authors found that training conducted at a subordinate level was more effective than the basic level in improving participants' subsequent ability to make same-different discriminations of birds. This shows that participants were better able to process and make high-level differentiations between the trained objects after undergoing a subordinate-level training.

The limitations of basic-level training tasks for perceptual training are particularly evident in an empirical study by Keyes et al. (2019). In an attempt to develop a motorcycle-specific perceptual training intervention, the authors examined the effectiveness of a perceptual training method based on visual search, which involved searching for motorcycles in a natural scene and reporting the number of motorcycles present. Unfortunately, no training benefits were found for participants' ability to detect motorcycles. However, the lack of training benefit is likely due to the fact that Keyes et al.'s (2019) perceptual training task was conducted at a basic level. In the training task, participant searched for motorcycles as a category and were not required to make within-category discriminations. As shown in the literature, training conducted at a basic level is not effective for developing perceptual expertise (Archambault et al., 1999; Scott et al., 2006; Tanaka et al., 2005).

In contrast, Crundall et al. (2017) proposed the use of a motorcycle-specific perceptual training task based on Pelmanism at a subordinate level. In their study, a single player version of Pelmanism was used and the goal of the game was to find all the matching pairs in the shortest time. Crucially, players matched motorcycle pairs by categorising and discriminating motorcycles at a subordinate level. This meant that players matched different types of motorcycles such as sports bikes and cruiser motorcycles, all of which could be differentiated based on front-on features such as headlights, handlebars and shape from the front view (see Figure 4.6).

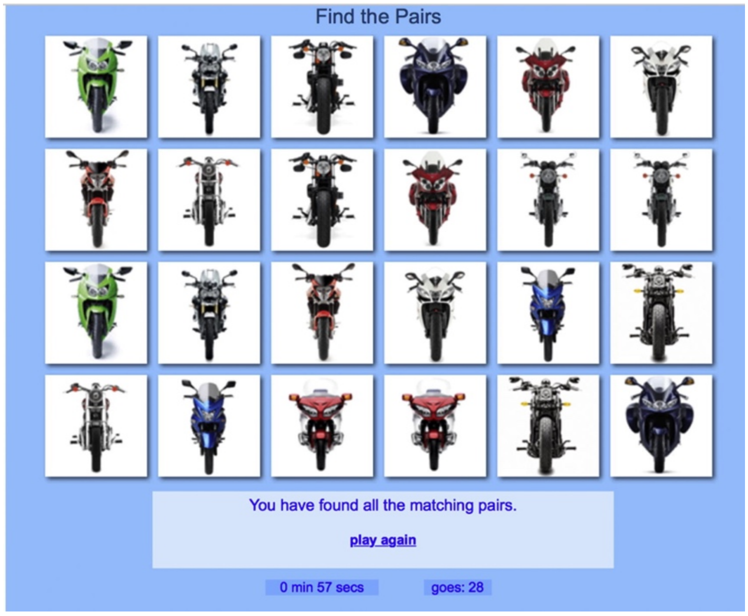

Figure 4.6: Pelmanism with motorcycle stimuli in Crundall et al.'s (2017) study. Permission to reproduce this has been granted by Elsevier.

Based on the literature reviewed, there is theoretical support for the subordinate-level Pelmanism task to improve drivers' perceptual skills for motorcycles because drivers are being trained to make within-category discriminations of motorcycles. Crundall et al.'s (2017) study reported that participants who played the motorcycle Pelmanism game subsequently improved in motorcycle detection accuracy on a T-junction task whereas those who played a fruit Pelmanism game in the control condition did not. The authors proposed that training players to discriminate between different types of motorcycles increased their sensitivity towards diagnostic features of motorcycles, thus increasing their ability to detect motorcycles. This is supported by evidence from another empirical study by van der Linden et al. (2014), who showed that training participants to discriminate between fish species subsequently increased their sensitivity to diagnostic features and reduced sensitivity to uninformative features. Although Crundall et al.'s (2017) study was a promising step towards developing an engaging and effective motorcycle-specific perceptual training tool, it is the only study to have investigated the use of Pelmanism as a method of perceptual training. The subsequent chapters in this thesis therefore aim to advance our understanding of a Pelmanism-based motorcycle training tool and further develop the version described in Crundall et al.'s (2017) study.

So far, a review of the theoretical bases of the game mechanic of Pelmanism suggests that a Pelmanism-based training tool has the potential to induce perceptual learning and develop perceptual expertise for motorcycles. Having reviewed the benefits of subordinate level discrimination and the matching process in Pelmanism, empirical research is needed to validate the motorcycle Pelmanism task. Given that subordinate-level categorisation is a hallmark of perceptual expertise (Bukach et al., 2010; Curby & Gauthier, 2009), it is expected that a motorcycle-specific Pelmanism task would be sensitive to players' familiarity with motorcycles because it requires players to match motorcycle pairs at a subordinate level. This is tested empirically and reported in the next chapter (Chapter 5).

4.3.2 Potential for using game elements in the Pelmanism-based training task

No game elements were used in Crundall et al.'s (2017) motorcycle Pelmanism task and there is scope to incorporate game elements as motivational features. Game elements are important because they facilitate learning and help achieve the intended training objective by keeping players engaged and motivated with the task. As mentioned in a previous section on the desired motivational features, game elements that appropriately challenge the player and are presented as specific goals while providing feedback of the player's progress contribute to the player's flow state (see Nakamura & Csikszentmihalyi, 2002). Difficulty levels and constructive competitions were identified as game elements that allow players to feel sufficiently challenged, while maintaining players' autonomy and intrinsic motivation (e.g., Bostan, & Öğüt, 2009; Fülöp, 2009). In turn, game incentives that reward players can be used to sustain players' motivation and persistence with the goals set by the challenges. Applying these game elements to the motorcycle Pelmanism task, clear and specific performance goals can be used to determine the level of difficulty in the game. For instance, there could be an easy and hard goal that challenge players to find all motorcycle pairs under a certain number of attempts, with more attempts representing the easy goal and fewer attempts representing the hard goal. Similarly, the game could also be played with time-based goals where longer times represent the easy goal and quicker times represent the hard goal. Crucially, players should be able to choose whether they want to aim for the easy or hard goal, in order to fulfil their need for perceived autonomy (see van Roy & Zaman, 2017). The performance goals could also be framed competitively to encourage constructive competitions, or they could offer rewards for attaining each goal. Research has yet to investigate the impact of such game elements in the context of the motorcycle Pelmanism task, and this thesis will be the first to evaluate the suitability of performance goals, competition and rewards in enhancing motivation and/or performance on the motorcycle Pelmanism task. The effects of game elements are examined and reported in Chapter 6.

4.4 Summary

The first half of this chapter reviewed the effectiveness of motorcycle awareness campaigns and existing hazard perception training interventions. It was noted that awareness campaigns were not effective in improving drivers' ability to spot motorcycles and existing hazard perception training interventions lacked emphasis on motorcycle hazards specifically. Another key limitation of existing hazard perception training interventions was the lack of appeal to drivers. Given that participation in hazard perception training is voluntary, the real-world impact of the training is very limited if drivers are not motivated to take part, regardless of how effective the training is in research settings. In light of this, a motorcycle-specific perceptual training method based on Pelmanism was explored in the second half of the chapter, drawing inspiration from the serious game literature. The conditions needed for players to achieve a state of flow during gameplay and the theoretical bases for a subordinate-level motorcycle perceptual training approach were discussed. Based on the evidence reviewed, training drivers to make subordinate-level discriminations of different motorcycles was found to be most promising in inducing perceptual learning and even developing perceptual expertise. To set up the appropriate conditions for flow state, several game elements were identified for the Pelmanism-based training task. These were specific goals framed as performance goals in the game, which could be competitive goals to encourage constructive social comparison or goals that offer rewards for accomplishing different goals. In the following chapter, the validity of the motorcycle Pelmanism task is examined by testing its sensitivity to dual drivers' and car drivers' performance on the task. Chapter 6 then examines the number of Pelmanism rounds and the game elements that make the motorcycle Pelmanism training task most effective and engaging. Based on the findings from these two chapters, an improved version of the motorcycle Pelmanism training task is then put together. The effectiveness of this motorcycle Pelmanism training task is evaluated in the final empirical study (Chapter 7), before a general discussion of the thesis is provided.

Chapter 5 – Validation of Pelmanism

Chapter overview

The potential of the motorcycle Pelmanism task to become a fully-fledged serious game to train perceptual expertise with motorcycles was reviewed in the previous chapter. The literature review indicated that the game mechanic of Pelmanism, which refers to the rules at the core of the game, should be able to deliver perceptual training effectively. This is because it requires players to make subordinate level discriminations when matching pairs of motorcycles from memory. As the motorcycle Pelmanism task is a relatively new approach, there is yet to be empirical evidence as to whether the game mechanic taps into relevant skills that develop with real world experience. This chapter therefore aims to test whether the Pelmanism task is sensitive to differences in motorcycle familiarity. Experiment 4 compares dual drivers' and car drivers' performance on the motorcycle Pelmanism task, by measuring both the number of attempts and the time taken to find all motorcycle pairs. As prior knowledge of motorcycles should facilitate discrimination between motorcycle features, it is expected that dual drivers, who are familiar with motorcycles, should perform better on the motorcycle Pelmanism task than car drivers. It is important to establish whether the motorcycle Pelmanism task is sensitive to differences in motorcycle familiarity because it ensures that the perceptual skills being trained reflect skills developed from real world exposure to motorcycles. Results from Experiment 4 indicated that dual drivers required fewer attempts than car drivers to find all motorcycle pairs, but surprisingly, this difference was not reflected in the time taken to complete the task. Experiment 5 was conducted with a different sample to see if results from Experiment 4 could be replicated and to further investigate why dual drivers' superiority in matching accuracy was not reflected in the overall time taken. The study differentiated overall time taken into three stages based on the basic actions that players perform during a game of Pelmanism. Similar to Experiment 4, dual drivers required fewer attempts than car drivers to find all matching motorcycle pairs, but in contrast to Experiment 4, dual drivers were also quicker overall than car drivers. As a result, there was no difference between driver groups in the average time taken at each stage.

5.1 Experiment 4: Sensitivity of the motorcycle Pelmanism task

5.1.1 Introduction

Although Pelmanism has been used as a memory training tool since the 1890s (Ennever, 2020), Crundall et al. (2017) were the first to explore its use in the context of motorcycle perception training. The authors proposed a motorcycle-specific perceptual training tool based on Pelmanism, using stimuli that varied at a subordinate level. In the previous chapter, the potential for a Pelmanism-based training task that capitalises on subordinate level matching of different motorcycle types was reviewed. Previous research on perceptual expertise indicates that training participants to make subordinate level categorisation, or within-category discriminations is crucial to improve perceptual skills, including detection and identification of objects from the trained category (Archambault et al., 1999; Scott et al., 2006; Tanaka et al., 2005). The motorcycle Pelmanism training task in Crundall et al.'s (2017) study was found to be effective in improving drivers' ability to detect motorcycles in a T-junction task. Drivers who completed the motorcycle Pelmanism task subsequently improved in motorcycle detection whereas drivers who completed a control Pelmanism task involving fruit pairs did not. Following this finding, more research is needed to continue developing and improving on the motorcycle Pelmanism training task.

Crundall et al.' s (2017) findings suggest that motorcycle Pelmanism task improves perceptual skills related to motorcycles. As such, we would expect that drivers who are familiar with motorcycles, such as dual drivers, would perform better than drivers who are not familiar with motorcycles in the absence of any training on the motorcycle Pelmanism task. However, this has not been tested. In Chapter 3, dual drivers were similarly expected to have a motorcycle-specific advantage because they are familiar with motorcycles, but no motorcycle-specific advantage across the three perceptual tasks was found. This highlights the need to empirically test any assumptions involving dual drivers' superiority in motorcycle-related tasks, such as the expectation that dual drivers would perform better than car drivers in a motorcycle-specific Pelmanism task. In addition, when developing a game-based training, it is important to establish construct validity, which is whether there are observable differences in the performance of skilled and unskilled players on the game (Graafland et al., 2012). This ensures that the skills being trained on the motorcycle Pelmanism task are the skills that players with high levels of knowledge and exposure to motorcycles possess. Access to motorcycle-specific information should produce an advantage for making subordinate level discriminations, and this advantage should be reflected by or translated into superior performance on the motorcycle Pelmanism task. In line with this, previous research on instructional tasks and serious games highlights the need for these tasks or games to be able to capture superior performance of experts because their aim is to teach skills that correspond to skills observed in experts (see Causer et al., 2014; Ericsson & Smith, 1991). According to perceptual expertise literature, experts have better memory abilities for objects in their domain of expertise (Herzmann & Curran, 2010; Rawson & Van Overschelde, 2008). Bukach et al. (2010) found that car experts were better at discriminating whether a car was the same or different to a previously presented car, compared to non-experts. This suggests that the car experts were not only able to discriminate between different types of cars presented but also remember the differences between car features. Drawing on these studies, dual drivers are expected to perform better on the motorcycle Pelmanism task because they have more prior knowledge and exposure to motorcycles than car drivers, which in turn facilitates the ability to discriminate between motorcycle pairs in the Pelmanism task. Therefore, empirical evidence is needed on whether the motorcycle Pelmanism task is sensitive to different levels of motorcycle familiarity.

Measures of performance on Pelmanism

Players' performance on the motorcycle Pelmanism task can be scored according to two different measures – time taken or number of attempts to match all the pairs in the grid. Time taken reflects the player's speed in matching whereas the number of attempts reflects the player's accuracy in matching. Previously, a study using Pelmanism to demonstrate adaptive memory found that participants made fewer mismatches when matching threatening images compared to unthreatening images (Wilson et al., 2011). This suggests that accuracy in the Pelmanism task may be influenced by top-down factors. However, it should be noted that the study only measured accuracy, and therefore was not a fair evaluation of the two performance measures. Although there is limited research involving Pelmanism specifically and how performance on the Pelmanism game may be influenced by familiarity with the matching pairs, previous research on memory accuracy suggests that experts have enhanced memory accuracy for objects within their domain of expertise. For example, as noted above, Herzmann and Curran (2011) found an accuracy advantage in car experts for recognising which of the car stimuli had been previously presented. Furthermore, Scolari et al. (2008) proposed that experts can maintain information in the working memory at a higher resolution than non-experts. Using a change detection task, they found an accuracy advantage for experts when changes occurred at the subordinate level but not at the basic level. The authors thus reasoned that experts encoded and recalled the stimuli at a higher level of detail than non-experts. The literature on experts' advantage in memory accuracy also suggests that experts are able to organise meaningful information based on domain-specific knowledge into chunks during encoding, by drawing on long-term memory representations such as schemata or mental templates for retrieval cues (Chase & Simon, 1973; Cowan, 2001; Gobet, 2005). It is possible that players with motorcycle-specific knowledge may be able to encode more features of a motorcycle by organising the features into chunks and processing how the features are in relation to each other, thus facilitating matching accuracy. This is similar to the notion that holistic processing underlies the expertise-related advantage in memory accuracy (see Curby et al., 2009). Therefore, players who are familiar with motorcycles may be more likely to encode and recall motorcycle stimuli accurately without revisiting previous cards, thus reducing the number of attempts needed to finish matching motorcycle pairs.

Previous literature on encoding durations suggests that time taken may also be a good measure of level of familiarity with motorcycles. This is based on research demonstrating that domain-specific knowledge can speed up consolidation of visual information into short-term memory. According to Xie and Zhang (2017), familiarity with particular stimuli increases the encoding speeds in the visual short-term memory. The authors found that participants encoded more familiar Pokemon characters within a limited amount of time than unfamiliar

Pokemon characters. This suggests that the rate of encoding increases with prior domain-specific knowledge. Using a backward masking design to manipulate the length of encoding duration, Curby and Gauthier (2009) found that car experts could process and encode cars more accurately at shorter durations than non-experts. Therefore, it is possible that players who are familiar with motorcycles may require less time to match motorcycle pairs because they can encode motorcycle stimuli at a faster rate than players who are not familiar with motorcycles. Based on the literature, motorcycle familiarity could benefit both the speed and accuracy on the motorcycle Pelmanism task. Therefore, both performance measures, time taken and number of attempts, will be recorded and analysed.

Baseline performance on Pelmanism

The ability to mentally hold visual images (i.e., visual representation) and store these images in the working memory is known to vary between individuals (see Reeder, 2017). In order to account for variance within participants when assessing performance on the Pelmanism task, a repeated measures design should be used whereby participants take part in two rounds of Pelmanism – one matching motorcycle pairs and the other as a control round. Previously, in Crundall et al.'s (2017) study, participants in the control condition were tasked to match fruit pairs. However, a limitation of using a fruit Pelmanism round as control is that participants may not be as engaged as participants in the motorcycle Pelmanism round. One of the reasons is that matching fruit pairs is not relevant to hazard perception. In addition, fruits can be easily identified at a subordinate level because they have commonly used labels such as apple or pear, whereas participants need to encode other details in order to discriminate between motorcycle pairs, such as yellow motorcycle with single headlight. As a result, there is a systematic difference in effort needed to encode and maintain fruit pairs compared to motorcycle pairs.

In light of this, a suitable control task that requires comparable levels of visual processing and encoding to matching motorcycle pairs is a pedestrian Pelmanism task. The reason is that pedestrians make up a recurring category of potential hazards that is equally familiar to both groups of drivers. Furthermore, participants can rely on details such as body features to discriminate between pedestrians (Rice et al., 2013), similar to discriminating between different motorcycles based on their features. As humans have been shown to be exceptionally good at processing and identifying people, face stimuli are commonly used in perceptual expertise research as a benchmark to compare object-related expertise against (e.g., Stein et al., 2016). Therefore, performance in the pedestrian Pelmanism task should function as a baseline measure of memory and feature discrimination ability while controlling for context relevance and engagement, because matching pedestrian pairs should be equally easy for all participants. This design should provide an accurate assessment of whether dual drivers exhibit superior performance specifically in matching motorcycle pairs compared to car drivers.

Current study

There is a need to establish the sensitivity of the motorcycle Pelmanism task to real world motorcycle familiarity in order to ensure that the content and underlying skills of the training taps into domain-specific information that drivers with motorcycle familiarity have access to. The aim of the current study is therefore to determine whether the motorcycle Pelmanism task is sensitive to players' familiarity with motorcycles, by comparing the performance of dual drivers and car drivers. Although a motorcycle-specific advantage for dual drivers was not found in the three specific tasks used in Chapter 3, there is still an inherent difference in levels of exposure to and knowledge of motorcycles between dual drivers and car drivers. Given that prior knowledge of the target category facilitates within-category discriminations (e.g., Herzmann & Curran, 2010; Xie & Zhang, 2017), dual drivers are likely to perform better than car drivers on the motorcycle Pelmanism task. Performance on the Pelmanism is measured in terms of both time taken and number of attempts because previous literature indicates that both speed and accuracy could be influenced by familiarity. A pedestrian Pelmanism round is used as a control to account for individual differences in memory performance. It is hypothesised that dual drivers would take a shorter time and fewer attempts on the motorcycle Pelmanism round than car drivers, but not on the pedestrian Pelmanism round.

5.1.2 Methods

Design

A 2x2 mixed design was used. The between-participants variable was driver group, which compared dual drivers against car drivers. The within-participants variable was the Pelmanism round – motorcycle or pedestrian (control). There were two outcome measures – number of attempts and time taken to find all 12 pairs. An addition third outcome variable, time taken per attempt, was also analysed. Time taken per attempt was calculated by dividing the time taken by number of attempts for each participant.

Participants

Sixty-nine participants took part, of whom 31 were dual drivers. Participants were recruited online via a paid participant panel and the university credit panel. Driver group status was determined based on participants' self-report on license(s) held. Three dual drivers and two car drivers did not report their car driving experience and were excluded from the analysis. The final sample size comprised 36 car drivers (15 females, 21 males) and 28 dual drivers (6 females, 22 males). The demographics of each driver group are shown in Table 5.1. Car driving experience and age were found to differ significantly between driver groups and were included as a covariate in the analysis.

Table 5.1: Demographics, driving history and habits of each driver group

	Car drivers	**Dual drivers**	***p***
Age	29.2 years (9.9)	34.7 years (11.4)	.046*
Frequency of playing memory games (mode)	Once or twice	Occasionally	-
Driving:			
Driving experience	9.5 years (8.2)	16.1 years (10.9)	.011*
Annual mileage	17583.3 miles (29282.2) [a]	7507.6 miles (4634.4)	.073
Hours driven per week	7.8 hours (5.1) [a]	10.8 hours (10.8)	.17

Riding:			
Riding experience	-	9.8 years (9.5)	-
Annual mileage	-	5928.0 miles (7714.3) [b]	-
Hours ridden per week	-	6.6 hours (7.7) [b]	-

[a] 8 missing cases, [b] 2 missing cases

*Below significance threshold of .05

Materials and stimuli

Three sets of Pelmanism stimuli were created – a practice round consisting of 12 car pairs, an experimental round consisting of 12 motorcycle pairs and a control round consisting of 12 pedestrian pairs. Twelve images each of front-facing cars, front-facing motorcycles and side-facing pedestrians were sourced from the internet to be used as target stimuli. Forty-eight empty roads were also sourced from the internet to be used as backgrounds for the target stimuli. Half of the empty roads were used as backgrounds for the practice round and the remaining half were used as backgrounds for both the experimental and control rounds. The 12 target stimuli were cropped out from the original image using the lasso tool in Adobe Photoshop and edited twice onto 24 backgrounds. Each pair of stimuli was created by editing one target stimulus onto two different backgrounds. This ensured that the target pairs could not be matched on the basis of their backgrounds. All backgrounds were cropped and resized to 1280 x 960 pixels. Similar to the use of different backgrounds within each pair, the size of the targets within each pair were not of the same size. The targets (i.e. car, motorcycle or pedestrian) were scaled to vary in size within each pair and edited onto the background according to the perspective of the road. On average, the targets were one third of the height of the background. This ensured that participants could not perform matches based on background and size. One target with one background made up one playing tile in the game, with the back of the playing tile displaying a question mark when the playing tile is face down (see Figure 5.1).

Figure 5.1: Example of a matched motorcycle pair across two playing tiles, and beneath them two unturned playing tiles.

The Pelmanism game was coded on Unity and hosted on an online platform (www.simmer.io). The tiles were set up in a 6x4 grid, starting face down. The 24 playing tiles were randomly allocated to a position at the start of each game. Tiles turned over to reveal the image when clicked. If two matching tiles were clicked, a sound effect was played to notify the player of a match and the tiles remained face up throughout the rest of the round. If the two selected tiles did not match, they would flip back over face down upon the next click. This process of selection and matching continued until all 12 pairs are found. A stopwatch timer and a counter were shown on the left side of the screen to record the time taken and number of attempts for each round. For every two tiles turned over, the attempt counter increased by one. Participants were also shown the number of matched pairs out of 12 on the left of the screen (see Figure 5.2 for a screenshot of the game). The timer was programmed to start only when the first tile in each round was clicked. Three Pelmanism rounds were created, one of which is a practice round with car stimuli and the other two with motorcycle and pedestrian stimuli for the actual Pelmanism game. A short demographics questionnaire was created at the end of the game to record participants' age, sex, driving (and riding) experience, as well as how often they play Pelmanism or similar memory games – never, once or twice, occasionally, or often.

Figure 5.2: Screenshot of the game

Procedure

This study was conducted online (http://bit.ly/PelmanismFullVer), and all participants gave informed consent prior to the start of the study. Participants were instructed to find all 12 matching pairs with as few attempts and as quickly as possible. The instructions also informed participants that the backgrounds and size of objects within each pair would be different, along with an image of a matching car pair to illustrate this. Before commencing the actual game, a practice round was available to ensure that participants understood the task. Participants could end the practice round at any point using a skip button once they felt that they understood the task. After the practice round, participants completed the motorcycle and pedestrian rounds in random order. At the end of the Pelmanism game, participants were asked to fill in the demographics questionnaire before being directed to their scores. Participants were then thanked and debriefed.

Statistical analysis

Data were analysed using the statistical programming language R (R Core Team, 2020). Data from the practice round were not analysed. The number of attempts and time taken to clear the round as well as time per attempt were analysed separately using linear mixed effect models from the lme4 package (Bates et al., 2015) and likelihood ratio tests as outlined in Chapter 2. The model accounted for variance within participants and included covariates of driving experience, order of rounds (motorcycle or pedestrian round first), and frequency of playing with memory games. As driving experience was a continuous variable, it was means centred such that the intercept represents the mean time taken or accuracy for the average participant when all other variables are held constant at baseline. The full models of each of the outcome variables are as follows:

Attempts / Time taken / Time per attempt ~ Driver group + Round + Driver group:Round + Car driving experience + Order of rounds + Memory games frequency + (1|Participant)

Reference/Baseline categories of fixed effects: Driver group - car driver, Round – pedestrian

The full model of each outcome variable performed better than a general linear model without random effects and the model fit comparisons are given in Appendix D.

5.1.3 Results

Number of attempts

The mean number of attempts to complete each Pelmanism round according to driver group is shown in Figure 5.3.

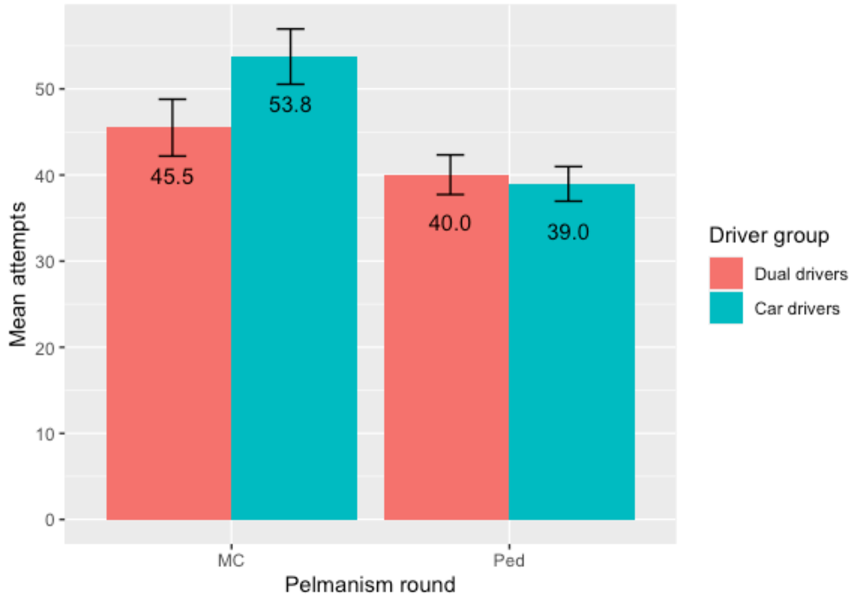

Figure 5.3: Mean number of attempts with +1/-1 SE bars to complete each Pelmanism round by driver group

Driver group was not a significant factor (χ^2 (1) = 0.59, p = .44; β = -2.75 [95% CI: -10.2, 4.71]), but the type of Pelmanism round was a significant factor (χ^2 (1) = 23.20, p < .001). Across both driver groups, participants required more attempts to match motorcycle pairs (M = 50.1, SD = 18.8) than pedestrian pairs (M = 39.4, SD = 12.0; β = 10.7 [95% CI: 6.71, 14.7]). The interaction between driver group and Pelmanism round was found to be significant (χ^2 (1) = 5.44, p = .020). Dual drivers made fewer attempts than experienced drivers in the motorcycle round but not the pedestrian round (t (62) = 2.34, p = .022; β = -9.31 [95% CI: -17.1, -1.53]; refer to Figure 5.3 for means).

Time taken

The mean time taken to complete each Pelmanism round according to driver group is shown in Figure 5.4.

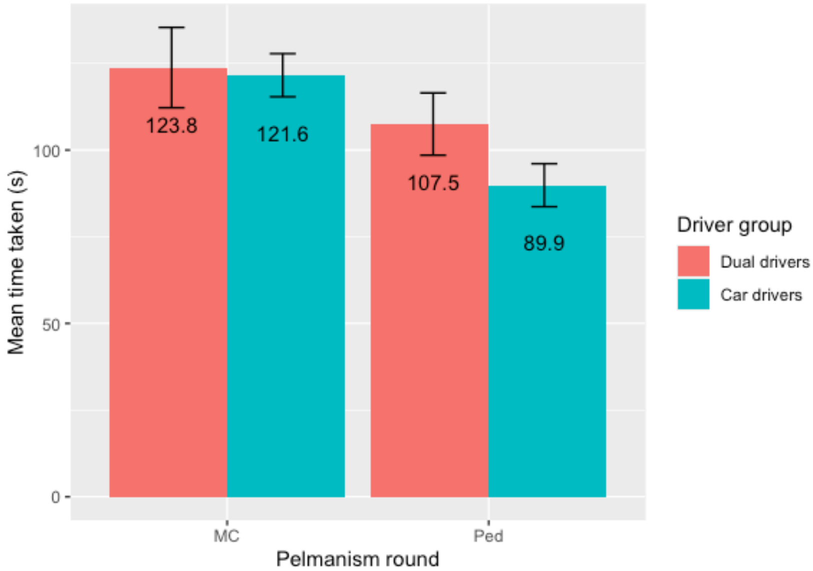

Figure 5.4: Mean time taken with +1/-1 SE bars to complete the motorcycle or pedestrian Pelmanism round by driver group

Driver group was not a significant factor (χ^2 (1) = 0.14, p = .70; β = 3.91 [95% CI: -17.3, 25.2]) but time taken to match all pairs differed significantly according to the Pelmanism round (χ^2 (1) = 20.30, p < .001). Across both driver groups, participants took longer to match all motorcycle pairs (M = 123.0s, SD = 48.8) than pedestrian pairs (M = 97.6s, SD = 42.6; β = 25.0 [95% CI: 14.9, 35.1]). The interaction between driver group and Pelmanism round was not significant (χ^2 (1) = 2.29, p = .13; β = -15.4 [95% CI: -35.6, 4.71]).

Time per attempt

The mean time per attempt according to driver group and Pelmanism round is shown in Figure 5.5.

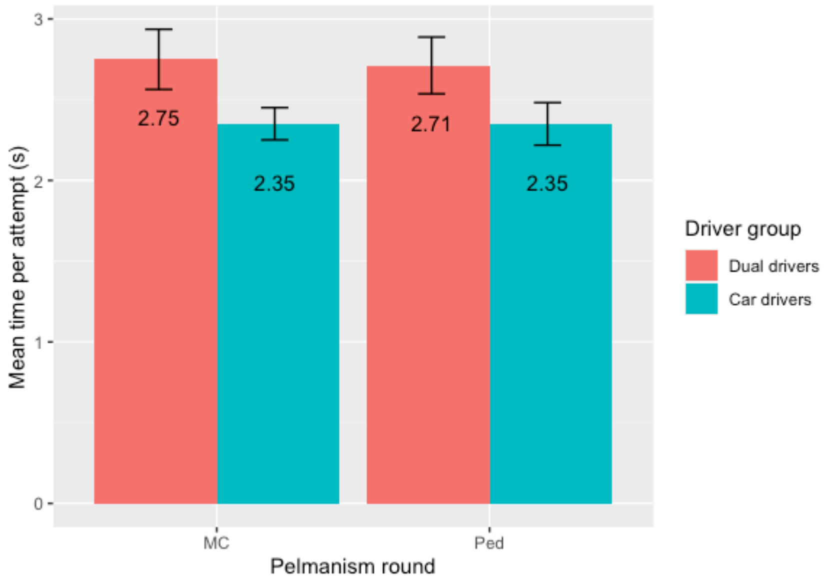

Figure 5.5: Mean time per attempt with +1/-1 SE bar by driver group and Pelmanism round

Driver group was not a significant factor (χ^2 (1) = 1.15, p = .28; β = 0.21 [95% CI: -0.20, 0.63]), and neither was the Pelmanism round (χ^2 (1) = 0.097, p = .75; β = 0.017 [95% CI: -0.089, 0.12]). The interaction between driver group and Pelmanism round was also not significant (χ^2 (1) = 0.12, p = .73; β = 0.037 [95% CI: -0.18, 0.25]).

5.1.4 Discussion

Analysis of the number of attempts needed to match all motorcycle pairs showed that dual drivers were more accurate in matching motorcycles compared to car drivers. The findings indicate that the motorcycle Pelmanism task is sensitive to level of familiarity with motorcycles; players who are familiar with motorcycles performed better in the motorcycle Pelmanism task than players who are not as familiar with motorcycles. The superior performance of dual drivers in the motorcycle Pelmanism task therefore provides some validation for the Pelmanism task and motorcycle stimuli because they were able to distinguish between the two driver groups. The current findings demonstrate that real world exposure and knowledge accumulated over the years is transferrable to the motorcycle Pelmanism task using the set of motorcycle pairs in the current study. Moreover, the finding that dual drivers displayed similar accuracy in matching motorcycle and pedestrians supports the notion that familiarity with the object category enhances performance in terms of number of attempts on the Pelmanism task.

Dual drivers' performance was superior to car drivers only in terms of number of attempts find all motorcycle pairs. This suggests that motorcycle-specific knowledge enhances performance on the motorcycle Pelmanism task by facilitating dual drivers' ability to encode and recall motorcycle features accurately. This is supported by research demonstrating that expertise enhances memory accuracy (Herzmann & Curran, 2011; Scolari, et al., 2008). In addition, accuracy in finding matched pairs on the Pelmanism task was also reported to reflect adaptive memory (Wilson et al., 2011). The number of attempts needed to finish matching all pairs is therefore a suitable measure to capture differences in memory performance. Dual drivers' ability to match motorcycle pairs accurately, as reflected by the number of attempts, indicates that motorcycle-specific knowledge enhances visual working memory for motorcycle stimuli. This is in line with previous research demonstrating the pre-existing knowledge improves encoding and subsequent recall or recognition (Bonasia et al., 2018; Lupyan, 2017).

Despite making fewer attempts in the motorcycle Pelmanism task, dual drivers were not faster at completing the round than car drivers. This pattern of results suggests that there might be a trade-off between accuracy and time taken for dual drivers to complete the round. However, the time taken per attempt as calculated by dividing the overall time taken for each participant by the number of attempts made does not tell us whether participants are spending the time during each attempt with one tile facing up or both tiles facing up. Therefore, there may be different ways in which time is spent even within each attempt during a game of Pelmanism. To further examine why dual drivers made fewer attempts but were not faster when matching motorcycle pairs, a follow up study is needed to disentangle the different stages of Pelmanism.

To summarise, Experiment 4 provided initial evidence that the motorcycle Pelmanism task is sensitive to the amount of real-world experience with motorcycles. The findings are promising because they validate the motorcycle Pelmanism task's ability to capture the superiority of dual drivers' knowledge and familiarity with motorcycles. An interesting pattern of results emerged whereby dual drivers required fewer attempts but were not quicker than car drivers to complete the motorcycle Pelmanism round. In light of this, a follow-up experiment was conducted to examine the different stages that occur during a game of Pelmanism. Experiment 5 attempts to replicate Experiment 4 by comparing the number of attempts and time taken by dual drivers and car drivers in the motorcycle Pelmanism round. However, Experiment 5 also uses an improved experimental paradigm to further investigate the trade-off between accuracy and speed observed in dual drivers in this study.

5.2 Experiment 5: Underlying mechanisms of dual driver advantage in motorcycle Pelmanism task

5.2.1 Introduction

Experiment 4 found that dual drivers needed fewer attempts but were not quicker to complete the Pelmanism task with motorcycle pairs compared to car drivers. The findings suggest that dual drivers may have traded speed for accuracy during the game. The time per attempt measure in Experiment 4 was obtained by dividing the overall time taken by the number of attempts. However, this does not provide a comprehensive understanding of how players allocate their time during each attempt because the different stages involved are not reflected appropriately when averaging the overall time taken. This experiment therefore addresses this by breaking down the overall time taken into different stages based on the actions that players make during the game. Examining time taken according to each stage can provide some insight into which stage(s) is/are influenced by familiarity and whether dual drivers allocate more time to certain stages during the Pelmanism in exchange for increased accuracy as compared to car drivers. For example, if dual drivers spend more time than car drivers when one tile is face up, this suggests that dual drivers spent increased time and effort remembering which tiles match.

The design of the matching process was such that two clicks were needed to complete an attempt and two scenarios were possible depending on whether the second click resulted in a successful match or not. For every attempt, the first click turns over a single tile and the second click turns over the second tile, which would result in either a match or mismatch with the first tile. The first stage occurs between the first and second click of each attempt, which will be referred to as the *match selection stage* because it refers to the selection of the second tile. During this stage, players reference their memory to determine if they recognise the stimulus and if so, which facedown tile is the matching stimulus. If the player does not recognise the stimulus or misremembers which tile is the matching stimulus, the second click is likely to result in a mismatch (although it is infrequently possible to match a new tile perfectly by chance). The mismatched tiles stay face up until another tile is clicked on, at which point the mismatched tiles return to be face down. The stage during which the mismatched tiles are face up will be referred to as the *mismatch stage,* and during this time, players are memorising the face-up tiles. If the second click turns over a matching tile, the two matching tiles remain visible throughout the rest of the task. Players then click on another single tile as the start of another attempt. Between clicking on the matching second tile and clicking on another single tile, there is no new information that is visible to prompt further cognitive processes and the time lag between these two clicks will be referred to as the *match delay stage.* The game of Pelmanism is therefore a cycle involving the *match selection stage*, *mismatch stage* and *match delay stage*. The player first turns over a single tile, processes it and tries to match it by turning over a second tile (*match selection stage*). If the two tiles do not match, they are visible until the next click (*mismatch stage*), which turns over another single tile for the *match selection stage*. If the two tiles match, the *match delay stage* occurs, and the player begins the next attempt by turning over a single tile again and triggers the *match selection stage.*

Admittedly, the time spent on the *match selection stage* is likely to vary throughout the game, similar to an inverted U-shape. At the beginning of the game, the time to select a match is likely to be short as players have no prior knowledge of the tiles and are unable to make informed matches. As the game progresses and players have accumulated knowledge of the tiles, the time spent on the *match selection stage* should increase as players have to reference their memory for matches. Towards the end of the game, as there is only a handful of unmatched tiles left, the time to select a match should decrease again. However, the variation in time spent on the *match selection stage* is not expected to differ across driver groups. Therefore, examining the average time spent on the *match selection stage*, the *mismatch stage* and *match delay stage* can still provide insight into whether dual drivers spend more time on certain Pelmanism stages compared to car drivers. Furthermore, as the experiment is conducted online, it is not feasible to record every click made by participants. Upon identifying the three stages that players experience during a game of Pelmanism, the following sections reviews the speed-accuracy trade-off at each of the stages and how dual drivers' familiarity with motorcycles may increase the time spent on each stage in exchange for increased accuracy.

Match selection stage

The *match selection stage* refers to the stage at which players try to find the tile that matches the first tile. The time spent on this stage is the time between the two clicks and represents the time taken to recall whether they recognise the stimulus on the first tile and which tile might be the matching stimulus. The ability to recall previously encoded information increases matching accuracy by reducing false matches, in which players recognise the stimulus but misremembers which tile is the matching stimulus. Previous research has demonstrated that recall accuracy increases for objects within the domain of expertise (e.g., Jackson & Raymond, 2008; Kalakoski & Saariluoma, 2001; Vogt & Magnussen, 2007). Therefore, it is possible that dual drivers are more accurate at recognising motorcycle stimuli than car drivers, resulting in fewer attempts needed. However, there is little evidence that increased recall accuracy occurs at the expense of time taken during the *match selection stage*. In other words, increased time spent selecting the second tile does not necessarily result in higher accuracy. This is because the majority of successful recalls occur within the initial recall period and the probability of recall decreases exponentially with time (see Rohrer & Wixted, 1994). Given that recall is dependent on successful encoding, it is also relevant to consider the stage during which encoding of motorcycle stimuli occurs – the *mismatch stage*.

Mismatch stage

The *mismatch stage* refers to cases where the two tiles turned over are not a match. As the two tiles remain face up until players click on another tile, the time between clicking on the second tile and clicking again represents the time spent on processing and encoding the mismatch. Successful encoding of the information can increase matching accuracy by reducing the number of redundant moves, which are moves that do not contribute to subsequent matches. For example, turning over a tile that had been previously turned over is a redundant move. Although revisiting tiles may sometimes be beneficial for reinforcing memories of previously turned over tiles, doing so frequently would increase the number of attempts to complete the Pelmanism task and thus reduce matching accuracy. As the *mismatch stage* reveals new or unexpected information, longer times spent during the *mismatch stage* increases the likelihood that information is being encoded. This is because encoding requires stimuli to be attended to and processed sufficiently in order to create long-lasting representations that are established in the visual working memory (see Bentin et al., 1998; Mangels et al., 2001). By the same logic, short times spent on the *mismatch stage* suggest that players may not be processing and encoding the information. Some players may choose to move on without encoding the unexpected stimulus, possibly because their working memory capacity is exceeded or to avoid interference of information already encoded in the working memory. For example, if a player knows that the matching tile is one of two possible tiles, he/she is likely to move on very quickly if the wrong tile was clicked on, in order to make a successful match in the next attempt. However, by doing so, they miss out on the opportunity to encode information about the wrong tile, meaning that they would have to revisit that tile before matching it, thus increasing the number of attempts. Therefore, short times spent on the *mismatch stage* may have a negative impact on matching accuracy and spending time to encode information during the *mismatch stage* may enhance matching accuracy.

Given that the motorcycle Pelmanism task requires players to perform matches at a subordinate level, encoding involves selective attention to and processing of higher-level details and motorcycle features. Based on previous research, players with motorcycle-specific knowledge should have access to subordinate-level information and the level of details stored in their visual working memory should correspondingly be higher than players without motorcycle-specific knowledge (see Harel, 2016; Scolari et al., 2008; Tanaka & Taylor, 1991). Therefore, dual drivers should be better able to encode higher level details and distinctive features of each motorcycles as compared to car drivers, who may not attend to motorcycle features to the same level of details. Consequently, when dual drivers encode high-level details and distinctive features of each motorcycle, more time may be required as compared to encoding low-level details. In addition, there is evidence that working memory capacity increases for objects within the domain of expertise (e.g., Chase & Simon, 1973; Curby et al., 2009). If dual drivers' visual working memory capacity allows them to store more motorcycle-related information than car drivers, dual drivers may spend more time whenever a mismatch occurs to encode the new information. In contrast, car drivers may move on quickly if the information is not directly relevant to the current stimulus they are matching, perhaps because the amount of motorcycle features to be encoded exceeds their working memory capacity. This means that car drivers have to revisit cards to encode them properly once their working memory capacity has been freed up, increasing the number of revisits and attempts as a result. Therefore, the time taken during the *mismatch stage* may vary between driver groups because dual drivers may spend more time to process the new or unexpected stimuli whereas car drivers may ignore the mismatch and carry on finding the intended match.

Match delay stage

The *match delay stage* occurs when two tiles turned over are a match. During this stage, no new information is visible and is equivalent to the start of the game where all tiles are facing down. The time lag between clicking on the matching tile and starting another attempt again represents the time taken to decide on which tile to match next and this decision could either be random, based on strategic locations on the board, or based on information accumulated from previous moves. Out of the three stages, the *match delay stage* is the least likely to affect matching accuracy directly because there is no information present to process or to cue recall as in the *mismatch stage* and *match selection stage* respectively. There is little theoretical basis and empirical evidence to conclude that matching accuracy could be increased at the expense of more time spent during this stage.

Current study

The aim of this study is to see if the same pattern of results in the previous study could be replicated, and to further investigate why dual drivers were more accurate but not quicker at matching motorcycle pairs than car drivers. Similar to the previous experiment, the current study compares the performance of dual drivers and car drivers on a motorcycle Pelmanism round and a control pedestrian Pelmanism round by measuring the number of attempts and overall time taken. Based on the findings from Experiment 4, dual drivers are hypothesised to require fewer attempts than car drivers to find all motorcycle pairs. An improvement to the previous study is the use of separate timers to record the time taken during each of the three stages mentioned in the introduction. Based on the literature reviewed and the fact that encoding directly impacts recall and matching accuracy, it is hypothesised that dual drivers would spend more time on the *mismatch stage* than car drivers during the motorcycle Pelmanism round.

One of the underlying reasons for this hypothesis was that dual drivers may have a greater working memory capacity for motorcycles than car drivers, and therefore be more likely to pay attention to mismatches and encode new or unexpected motorcycle stimuli whenever possible. However, it is also possible that differences in strategy rather than working memory capacity underlie any differences between dual drivers and car drivers. Dual drivers may prioritise accuracy over speed because they are compelled to identify the different motorcycles correctly. In contrast, car drivers may prioritise speed over accuracy to overcome their lack of familiarity with motorcycles, thus leading to a fast random clicking strategy. To address this concern of whether there are differences in working memory capacity or strategies, half of the participants in the current study would be subject to a three-second delay during the *mismatch stage*. In the normal (i.e., non-enforced) condition, the duration of the *mismatch stage* is determined by when players click on another single tile. In the enforced condition, the mismatched tiles would remain face up for a fixed duration of three seconds and the *mismatch* *stage* cannot be terminated by players. This is to prevent players from using a speed-based strategy and to encourage all players to process the mismatches whenever they occur. Therefore, if dual drivers' advantage in motorcycle matching accuracy is due to differences in strategy, the three-second delay should enforce car drivers to process mismatches and thus enhance matching accuracy by reducing redundant moves. As a result, an interaction would be observed whereby car drivers become more accurate in matching motorcycle pairs when subject to a three-second delay compared to the non-enforced condition, but there would be no difference for dual drivers. However, if dual drivers' advantage in matching accuracy is due to differences in familiarity and working memory capacity, car drivers should not be affected by the three-second delay and no interaction would be observed. In summary, the hypotheses are:

1) Dual drivers would require fewer attempts than car drivers in the motorcycle Pelmanism round, but not in the pedestrian Pelmanism round.
2) Dual drivers' accuracy advantage is at the expense of spending more time during the *mismatch stage* than car drivers.
3) If dual drivers' accuracy advantage is due to strategy, car drivers' accuracy in the motorcycle Pelmanism round should be higher in the enforced condition than in the non-enforced condition, but not for dual drivers.

5.2.2 Methodology

Design

A 2x2x2 mixed design was used. The first between-participants variable was driver group, which compared dual drivers and car drivers. The within-participants variable was the Pelmanism round – motorcycle or pedestrian (control). A second between-participants variable was the presence of enforced delay – enforced or non-enforced condition. Five outcome variables were measured and analysed separately – number of attempts, overall time and time spent during each of the three stages. These will be referred to as *match selection time*, *mismatch time*, and *match delay* for the purpose of the analyses. Figures 5.6 and 5.7 illustrate the three stages during the enforced condition and non-enforced condition respectively.

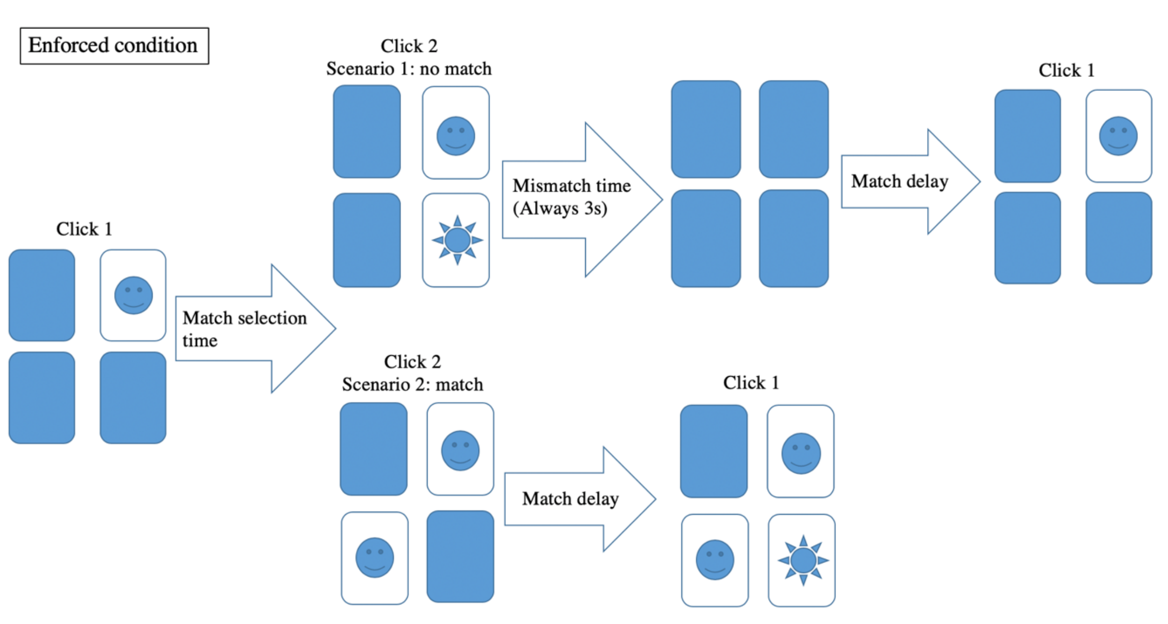

Figure 5.6: Timer design for enforced condition

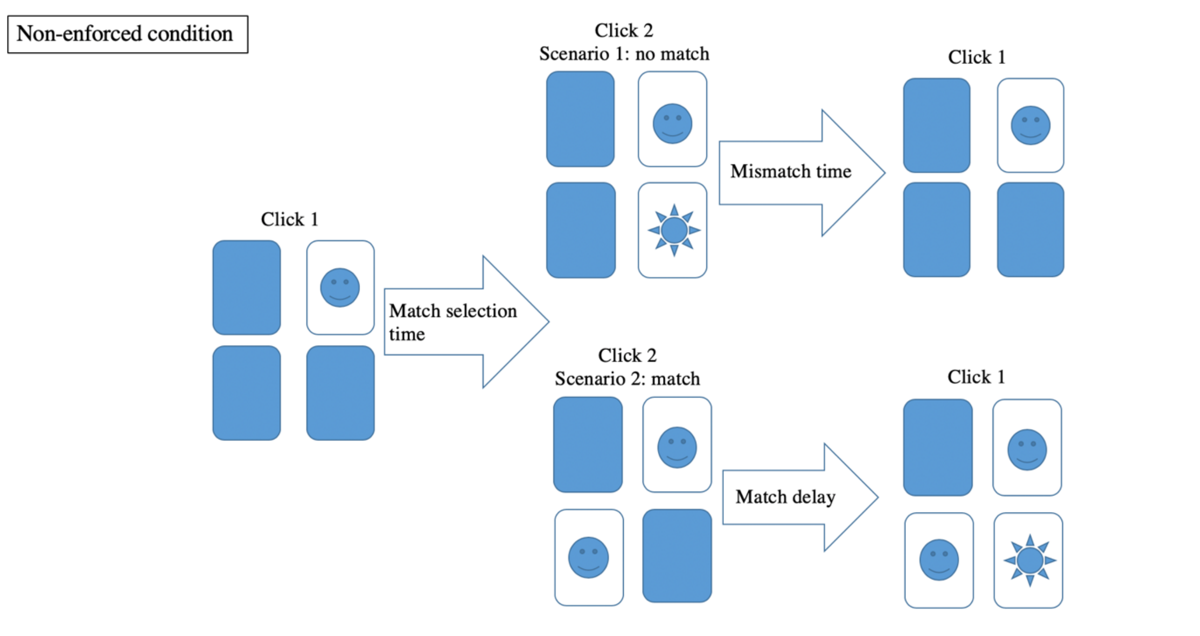

Figure 5.7: Timer design for non-enforced condition

Participants

Eighty-seven participants took part, of whom 41 were dual drivers. Participants in each driver group were randomly allocated to the enforced or non-enforced conditions. Participants were recruited via a paid participant panel known as Prolific and the link was also advertised to dual drivers via motorcycling groups. Driver group status was determined based on participants' self-report on license(s) held. One motorcyclist did not report having a driving license and two dual drivers did not report their driving and riding experience, therefore they were excluded from the analysis. The final sample consisted of 43 car drivers (14 males), of whom 22 were allocated the enforced condition, and 41 dual drivers (8 females), of whom 20 were allocated the enforced condition. The demographics of each driver group are shown in Table 5.2. Driving experience was significantly different between driver groups and was included as a covariate in the analysis.

Table 5.2: Demographics of each driver group

	Car drivers	**Dual drivers**	***p***
Age	46.7 years (13.2)	50.0 years (13.8)	.28
Frequency of playing memory games (mode)	Occasionally	Never	-
Driving:			
Driving experience	23.3 years (12.1)	31.0 years (14.6)	.011*
Annual mileage	6667.4 miles (3924.0)	8718.4 miles (5868.0)	.065
Hours driven per week	7.4 hours (6.1)	7.6 hours (6.3) [a]	.91
Riding:			
Riding experience	-	25.6 years (16.7)	-
Annual mileage	-	8082.8 miles (6690.7)	-
Hours ridden per week	-	9.3 hours (15.7)	-

*Below significance threshold of .05

[a] 1 missing case

Materials and stimuli

The same materials and stimuli from Experiment 4 were used to create the motorcycle and pedestrian rounds as well as the demographics questionnaire, but three changes were made to the Pelmanism game. First, the practice round was shortened to only three pairs and secondly, an additional enforced delay condition was programmed and allocated randomly to participants at the start of the study. In the enforced condition, a mismatch would trigger the cursor to be inactive for three seconds while the mismatched tiles remained visible. After three seconds, the cursor would become active and the mismatched tiles would flip back over facedown. In the non-enforced condition, the mismatched tiles remained visible until players click on another single tile again, similar to the version in Experiment 4. There was no fixed duration of three seconds and players were able to look at the mismatches for as long or as short as they would like. The final change was the addition of three timers to record the time spent during the match selection, mismatch and match delay stages separately (refer to Figures 5.6 and 5.7). Each of these timers recorded the running total of the time taken during each stage. The timers ran in the background and were not shown to participants. Participants saw only the overall timer during the game as with the previous version. The *match selection time* was recorded from the first click to the second click of each attempt and the *mismatch time* was recorded from when the mismatched tiles were visible until they returned face down. For the enforced condition, the *mismatch time* was always fixed at three seconds. Lastly, the *match delay* was recorded from clicking on a matching second tile and clicking on another single tile as the beginning of another attempt. For the enforced condition, the *match delay* also included the time from when the enforced delay ended (i.e., when mismatch tiles return facedown) until players clicked on another single tile again because there no information is visible at this point.

Procedure

The experimental procedure was similar to Experiment 4. The study was conducted online (https://bit.ly/Pelmanism2), and all participants gave informed consent prior to the start of the study. Participants were then randomly allocated to the enforced condition or non-enforced condition. The same instructions from Experiment 4 were shown but in the enforced condition, participants were additionally informed that the tiles would stay faceup for three seconds if they were mismatched (see Figure 5.8). A practice round was available using three pairs of car stimuli, to ensure that participants understood the task. Participants could end the practice round at any point using a skip button once they felt that they understood the task. After the practice round, participants completed the motorcycle and pedestrian rounds in random order. At the end of the Pelmanism game, participants were asked to fill in the demographics questionnaire before being directed to their scores. Participants were then thanked and debriefed.

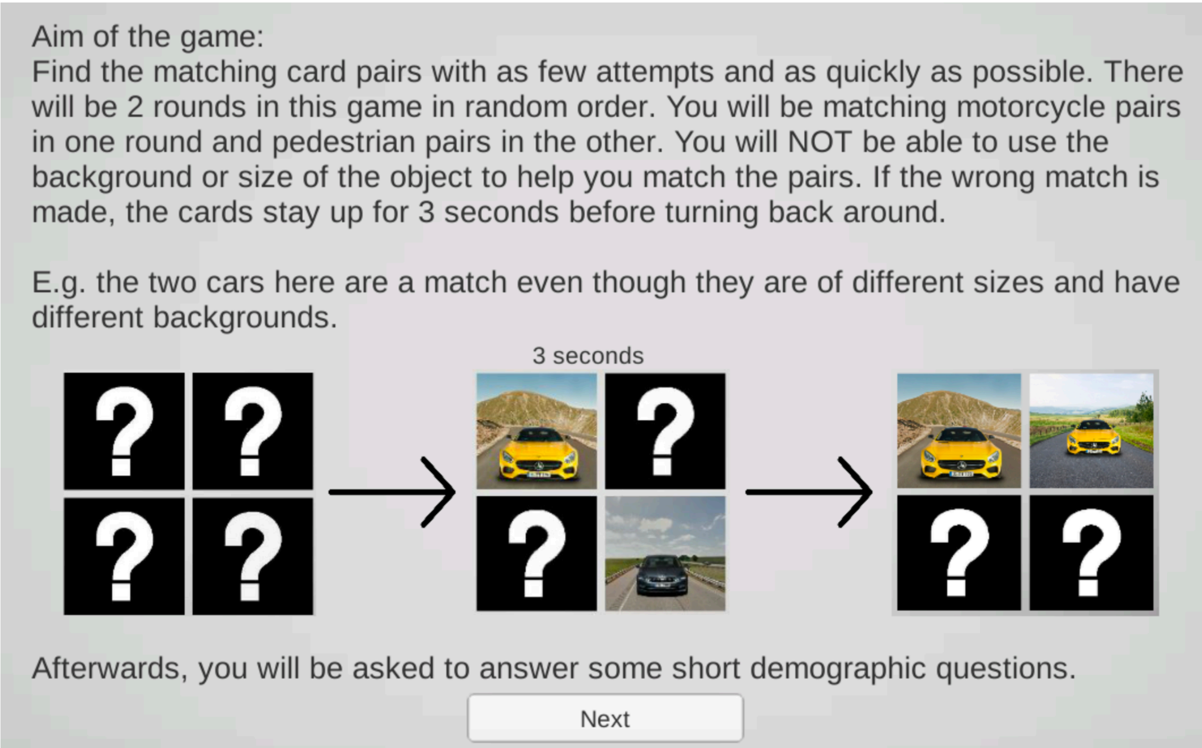

Figure 5.8: Instructions for enforced delay condition

Analysis

Data were analysed using the statistical programming language R (R Core Team, 2020). The five outcome variables, namely the number of attempts, overall time taken and the average time spent on each of the three stages, were analysed separately. Linear mixed effect models and likelihood ratio teste were used to analyse each of the dependent variables (refer to

Chapter 2). The model accounted for random variance within participants and included driving experience, order of rounds and frequency of playing memory games as fixed effect covariates. Driving experience was means centred. The full models are similar for all four outcome variables and are as follows:

Outcome ~ Driver group + Round + Presence of enforced delay + Driver group:Round: Presence of enforced delay + Driving experience + Order of rounds + Memory games frequency + (1|Participant)

Reference/Baseline categories of fixed effects: Driver group - car driver, Round – pedestrian, Presence of enforced delay – non-enforced

The full model for each outcome variable performed significantly than the respective general linear model without random effects (refer to Appendix E for model fit comparison).

5.2.3 Results

Number of attempts

The mean number of attempts to complete each Pelmanism round by each driver group in the enforced or non-enforced condition is shown in Figure 5.9.

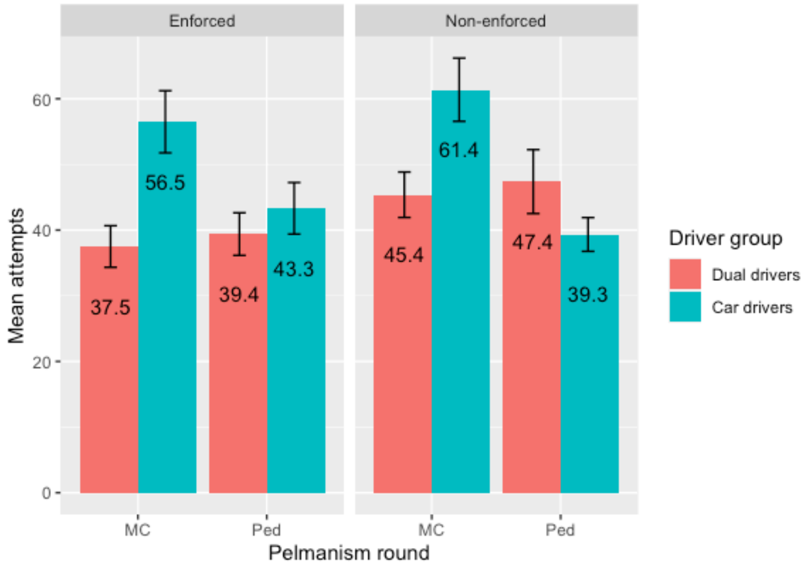

Figure 5.9: Mean number of attempts with +1/-1 SE by Pelmanism round, driver group and presence of enforced delay

The likelihood ratio test statistics are presented in Table 5.3. Likelihood ratio tests found that driver group was a significant factor as dual drivers (M = 42.5, SD = 17.3) made fewer attempts than car drivers (M = 50.1, SD = 20.9). Type of Pelmanism round was also a significant factor. Participants made more attempts in the motorcycle round (M = 50.4, SD = 20.9) than the pedestrian round (M = 42.4, SD = 17.3). However, presence of enforced delay was not a significant factor.

Table 5.3: Summary of main effects analyses for number of attempts (below significance threshold of .05)*

Main effect	**Df**	χ^2	p	**β [95% CI]**
Driver group	1	9.97	.0016*	-11.0 [-17.9, -4.02]
Round	1	12.72	< .001*	8.01 [3.75, 12.3]
Presence of enforced delay	1	2.88	.090	-5.62 [-12.4, 1.14]

Testing the two-way and three-way interactions with the drop1 function, the interaction between driver group and Pelmanism round was significant. Dual drivers (M = 41.5, SD = 15.4) made fewer attempts than car drivers (M = 58.9, SD = 22.0) in the motorcycle round, but not in the pedestrian round (t (82) = 5.10, $p < .001$). In the pedestrian round, there was no difference in the number of attempts between dual drivers (M = 43.5, SD = 19.1) and car drivers (M = 41.4, SD = 15.5). No other interactions were significant (see Table 5.4).

Table 5.4: Summary of interaction analyses for number of attempts (below significance threshold of .05)*

Interaction	**Df**	χ^2	p	**β [95% CI]**
Driver group x Round	1	23.73	< .001*	-19.6 [-27.0, -12.1]
Driver group x Presence of enforced delay	1	1.41	.23	-7.63 [-20.9, 5.62]
Round x Presence of enforced delay	1	1.43	.23	-4.49 [-11.9, 2.97]
Driver group x Round x Presence of enforced delay	1	1.45	.23	8.97 [-5.92, 23.8]

Overall time taken

The mean overall time taken to complete each Pelmanism round according to each driver group in the enforced or non-enforced condition is shown in Figure 5.10.

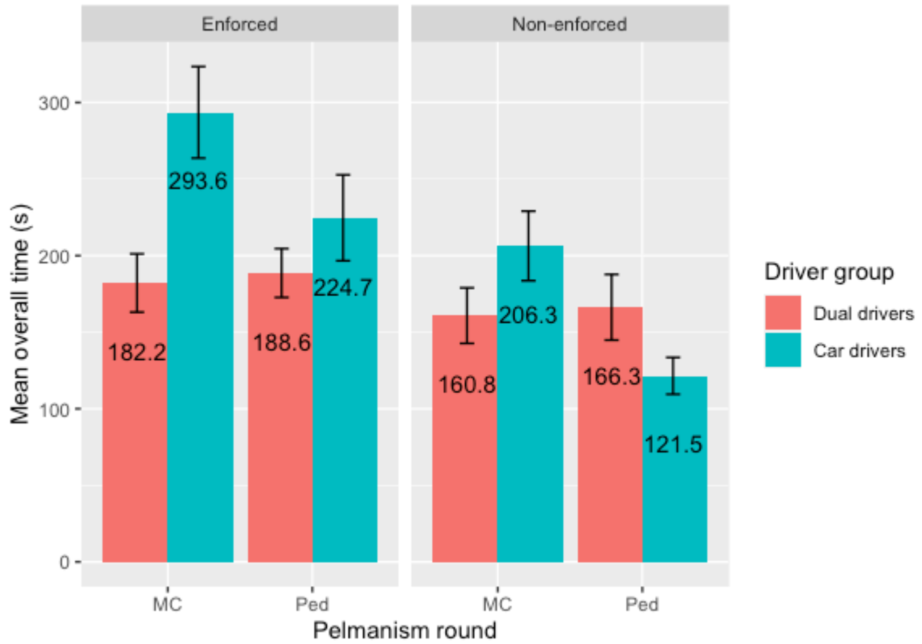

Figure 5.10: Mean overall time taken with +1/-1 SE by Pelmanism round, driver group and presence of enforced delay

Likelihood ratio test statistics are presented in Table 5.5. Driver group was a significant factor, with car drivers (M = 213.0s, SD = 127.0) taking more time than dual drivers (M = 174.0s, SD = 84.3). The type of Pelmanism round was also a significant factor; participants took more time to complete the motorcycle round (M = 212.0s, SD = 116.0) than the pedestrian round (M = 176.0s, SD = 100.0). Unsurprisingly, the presence of enforced delay was a significant factor; participants in the enforced condition took more time to complete the round (M = 224.0s, SD = 119.0) than participants in the non-enforced condition (M = 164.0s, SD = 90.8).

Table 5.5: Summary of main effects analyses for overall time taken (below significance threshold of .05)*

Main effect	**Df**	χ^2	p	**β [95% CI]**
Driver group	1	13.96	< .001*	-67.7 [-103.0, -31.9]
Round	1	10.85	< .001*	36.3

				[15.3, 57.4]
Presence of enforced delay	1	8.24	.004*	49.8 [28.0, 65.5]

Likelihood ratio test statistics on two-way and three-way interactions are presented in Table 5.6. The two-way interaction between driver group and Pelmanism round was significant. In the motorcycle round, dual drivers (M = 171.0s, SD = 83.8) were quicker than car drivers (M = 251.0s, SD = 130.0), but not in the pedestrian round (t (81) = 4.20, $p < .001$). In the pedestrian round, there was no difference in overall time taken between dual drivers (M = 177.0s, SD = 85.8) and car drivers (M = 174.0s, SD = 113.0). The two-way interaction between driver group and presence of enforced condition was also significant. Dual drivers (M = 185.0s, SD = 77.5) were quicker than car drivers (M = 259.0s, SD = 139.0) in the enforced condition, but not in the non-enforced condition (t (75) = 2.07, $p = .042$). In the non-enforced condition, there was no difference in overall time taken between dual drivers (M = 164.0s, SD = 89.9) and car drivers (M = 164.0s, SD = 92.8). No other interactions were significant.

Table 5.6: Summary of interaction analyses for overall time taken

Interaction	**Df**	χ^2	p	**β [95% CI]**
Driver group x Round	1	16.56	< .001*	-82.8 [-121.0, -44.2]
Driver group x Presence of enforced delay	1	4.67	.03*	-70.8 [-138.0, -3.81]
Round x Presence of enforced delay	1	0.20	.65	-8.67 [-47.3, 29.9]
Driver group x Round x Presence of enforced delay	1	0.15	.70	14.9 [-62.8, 92.6]

Match selection time

The mean match selection time during the motorcycle and pedestrian rounds according to each driver group in the enforced or non-enforced condition is shown in Figure 5.11.

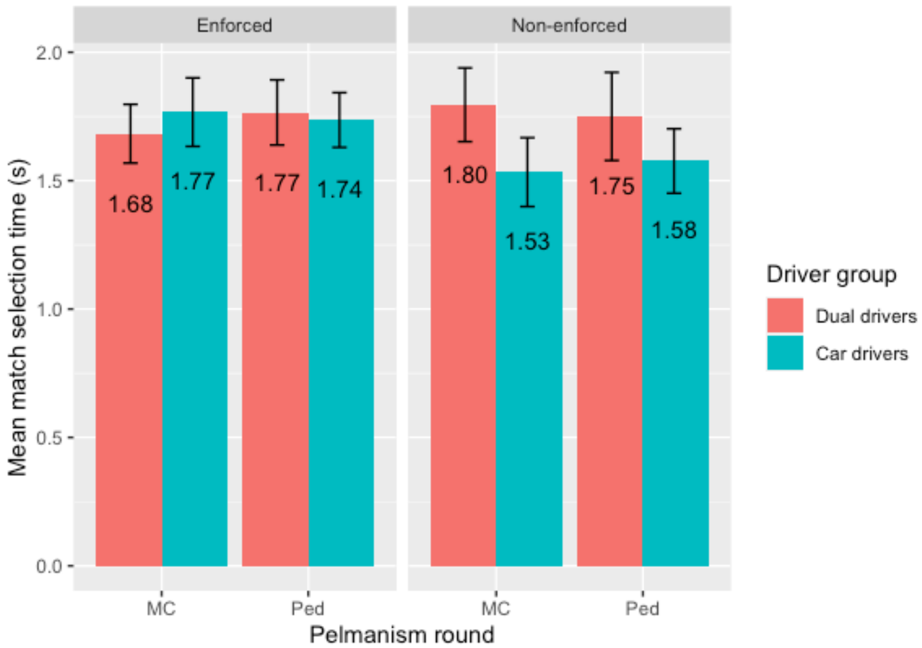

Figure 5.11: Mean time spent on the match selection stage with +1/-1 SE by Pelmanism round, driver group and presence of enforced delay

Likelihood ratio tests found no main effects of driver group, round nor presence of enforced delay (see Table 5.7). The two-way and three-way interaction between driver group, round and condition were also tested using likelihood ratio tests. None of the interactions were significant. The likelihood ratio test statistics are presented in Table 5.7.

Table 5.7: Summary of main effects and interaction analyses for match selection time

Main effect	**Df**	χ^2	p	**β [95% CI]**
Driver group	1	0.78	.38	-0.10 [-0.32, 0.13]
Round	1	0.079	.78	-0.011

				[-0.090, 0.068]
Presence of enforced delay	1	0.16	.69	0.042 [-0.18, 0.26]
Interaction				
Driver group x Round	1	0.024	.88	-0.012 [-0.17, 0.15]
Driver group x Presence of enforced delay	1	1.05	.30	-0.21 [-0.64, 0.22]
Round x Presence of enforced delay	1	0.096	.76	-0.025 [-0.18, 0.14]
Driver group x Round x Presence of enforced delay	1	1.62	.20	-0.20 [-0.52, 0.12]

Mismatch time

The mean time spent during the mismatch stage according to driver group and Pelmanism round is shown in Figure 5.12. Only data from the non-enforced condition were analysed because the mean time taken during mismatch stage was always three seconds for the enforced condition.

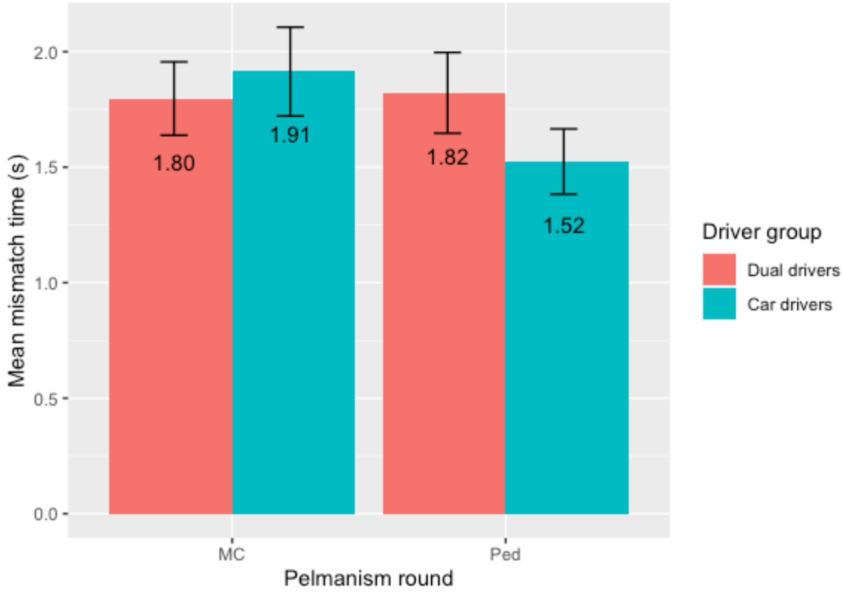

Figure 5.12: Mean time spent on the mismatch stage with +1/-1 SE by Pelmanism round, driver group and presence of enforced delay

Likelihood ratio tests found no significant main effects of driver group and round. Although the means suggest that car drivers spent more time during motorcycle mismatches than pedestrian mismatches relative to dual drivers, this interaction did not reach significance (see Table 5.8).

Table 5.8: Summary of main effects and interaction analyses for mismatch time

Main effect	**Df**	χ^2	*p*	**β [95% CI]**
Driver group	1	0.78	.38	-0.058 [-0.46, 0.34]
Round	1	0.079	.78	0.18 [-0.037, 0.40]
Interaction				
Driver group x Round	1	3.68	.055	-0.41 [-0.84, 0.010]

Match delay

The mean match delay for each Pelmanism round and driver group in the enforced or non-enforced condition is shown in Figure 5.13.

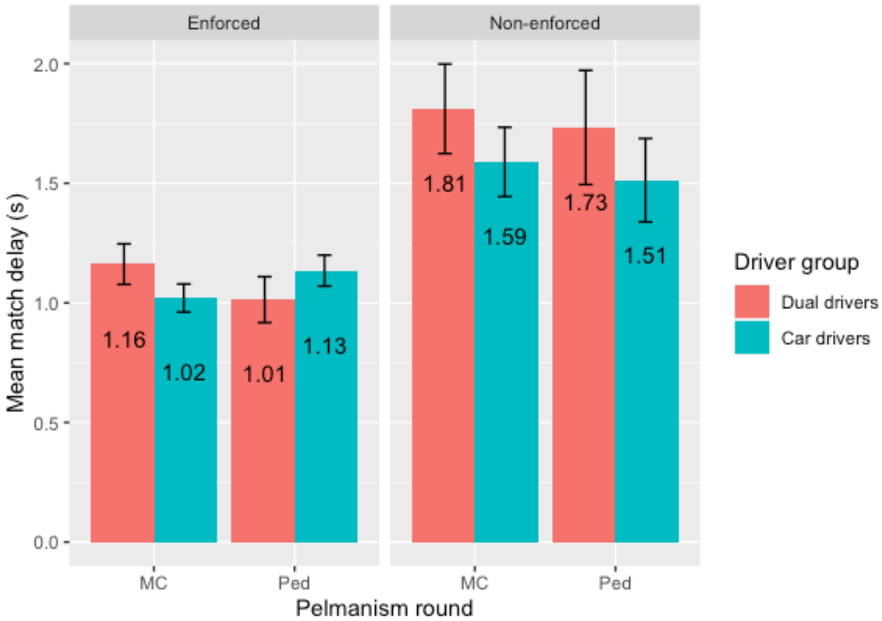

Figure 5.13: Mean time spent on match delay stage with +1/-1 SE by Pelmanism round, driver group and presence of enforced delay

Likelihood ratio tests found no significant main effect of driver group and round, except for condition (see Table 5.9). Participants in the enforced condition spent significantly less time during the match delay (M = 1.08s, SD = 0.35) than participants in the non-enforced condition (M = 1.66s, SD = 0.86).

Table 5.9: Summary of main effects analyses for match delay (below significance threshold of .05)*

Main effect	Df	χ^2	*p*	β [95% CI]
Driver group	1	0.99	.32	-0.11 [-0.34, 0.12]
Round	1	0.83	.36	0.044 [-0.051, 0.14]

Condition	1	30.0	< .001*	-0.65 [-0.88, -0.43]

The two-way and three-way interaction between driver group, round and condition were also tested using likelihood ratio tests. None of the interactions were significant (see Table 5.10).

Table 5.10: Summary of interaction analyses for match delay

Interaction	**Df**	**χ^2**	***p***	**β [95% CI]**
Driver group x Round	1	1.90	.17	0.13 [-0.058, 0.32]
Driver group x Condition	1	0.44	.51	-0.20 [-0.64, 0.24]
Round x Condition	1	0.90	.34	-0.063 [-0.25, 0.13]
Driver group x Round x Condition	1	1.91	.17	0.26 [-0.12, 0.64]

5.2.4 Discussion

The present study compared the performance of dual drivers and car drivers on the motorcycle and pedestrian Pelmanism tasks by measuring the number of attempts and overall time taken to complete each round. The overall time taken was further broken down according to the different stages of Pelmanism, namely *match selection stage*, *mismatch stage* and *match delay stage*. Furthermore, the effect of a three-second enforced delay during the *mismatch stage* was examined. Using a different group of participants, the present study corroborated the findings in Experiment 4 that dual drivers made fewer attempts than car drivers in the motorcycle Pelmanism round but not the pedestrian round. This provides further evidence that the motorcycle Pelmanism task is sensitive to players' level of familiarity with motorcycles and is in line with previous research on perceptual expertise demonstrating that experts have increased memory accuracy for objects within their domain of expertise (Herzmann & Curran, 2011; Scolari, et al., 2008).

An unexpected finding was that dual drivers were also quicker than car drivers to complete the motorcycle Pelmanism round, but not the pedestrian Pelmanism round. This finding was unexpected because Experiment 4 did not show a difference between dual drivers and car drivers in terms of the overall time taken. It is possible that dual drivers in the present study had both speed and accuracy advantages in the motorcycle Pelmanism round. However, the analysis also revealed a two-way interaction between driver group and presence of enforced delay for overall time taken. This raises the possibility that dual drivers' speed advantage could have been driven by the enforced delay because half of the dual drivers were subject to the three-second delay, in which they were overall faster than car drivers. Furthermore, it is possible that dual drivers' advantage in overall time taken for the motorcycle Pelmanism round was in part due to their advantage in matching accuracy, and the effect of accuracy on overall time taken may have been more pronounced with the enforced delay as compared to the non-enforced delay condition. As dual drivers made fewer attempts than car drivers, dual drivers would have encountered fewer counts of mismatches and enforced delay, therefore they were less affected by the presence of enforced delay than car drivers. In contrast, although car drivers could get away with making more attempts without losing too much time in the non-enforced condition, they were penalised by the three-second delay in the enforced condition such that car drivers were significantly more time overall than dual drivers. However, as there was no three-way interaction between driver group, Pelmanism round and presence of enforced delay, there is insufficient evidence to conclude that dual drivers' advantage for overall time taken was due to the presence of enforcement.

The present study also investigated whether there were driver group differences in how time was allocated within each Pelmanism stage. No driver group differences were found in each of the stages. The hypothesis that dual drivers would spend more time during the *mismatch stage* than car drivers was therefore not supported. This could be attributed to the absence of speed-accuracy trade-off in the present study. As dual drivers were quicker overall and more accurate than car drivers in the motorcycle Pelmanism round, the average time spent on each Pelmanism stage should not be affected because accuracy did not increase at the expense of time as in Experiment 4. Therefore, it is reasonable that the average time spent on each stage did not differ between driver groups.

The effect of the three-second enforced delay on participants' accuracy (i.e., number of attempts) was also investigated and no effect was found. There was also no interaction between driver group and the presence of enforced delay for the number of attempts. As car drivers did not benefit from the additional time provided during the mismatch stage to process new or unexpected stimuli, this rules out the possibility that dual drivers' accuracy advantage in the motorcycle Pelmanism round was due strategic differences at the *mismatch stage*. Therefore, it is possible that dual drivers require fewer attempts than car drivers because their working memory capacity allows them to encode new information during mismatches whenever possible, thus reducing the number of redundant moves. An unanticipated finding resulting from the presence of the three-second delay was that the match delay time was shorter in the enforced condition than the non-enforced condition. A potential reason is that participants in the enforced condition used a portion of the three-second delay to decide which tile to click on next, thus reducing the match delay time. Therefore, it is likely that the average match delay time differed systematically between enforced and non-enforced conditions. As the classification of the Pelmanism into distinct stages proposed in the current study is novel, future studies should refine the distinction of stages and consider the effect of an enforced delay on the match delay time.

In conclusion, dual drivers exhibited superior matching accuracy by requiring fewer attempts in the motorcycle Pelmanism task, as found in Experiment 4. Dual drivers were also quicker to match all motorcycle pairs compared to car drivers, thus no driver group difference in the average time spent on the *match selection stage*, *mismatch stage* and *match delay stage* were found. However, the possibility that the dual driver advantage in overall time taken was driven by dual drivers benefitting from the three-second delay and car drivers being penalised by the three-second delay for making more attempts than dual drivers was raised. With regards to the number of attempts measure, the enforcement of a three-second delay during the *mismatch stage* did not benefit car drivers' accuracy on the motorcycle Pelmanism task, suggesting that the accuracy advantage demonstrated by dual drivers was not due to differences in strategies between car drivers and dual drivers. Instead, it is likely that dual drivers are more accurate at matching motorcycle pairs because of their familiarity with motorcycles.

5.3 General discussion

This chapter reported two experiments that assessed the sensitivity of the motorcycle Pelmanism task to players' level of knowledge and experience with motorcycles. The findings are promising in that the motorcycle Pelmanism task and the stimuli used were able to distinguish between players who are more familiar with motorcycles, dual drivers, from players who are not as familiar with motorcycles, car drivers. Across both experiments, dual drivers were found to be more accurate at matching motorcycle pairs from memory than car drivers. Given that dual drivers have more exposure to and experience with motorcycles, it is reasonable that they would perform better on a task that requires detailed information of motorcycles as compared to car drivers. This is in line with previous research demonstrating that domain-specific knowledge enhances the ability to make subordinate level categorisations and also remember objects within that domain (Bukach et al., 2010; Herzmann & Curran, 2010).

In Experiment 4, there was no difference between driver groups in the overall time taken to complete the motorcycle Pelmanism task. In contrast, Experiment 5 found that dual drivers were faster than car drivers to complete the motorcycle Pelmanism task. However, this finding must be interpreted with caution due to other interactions being present. Specifically, the presence of a dual driver advantage in overall time taken in Experiment 5 but not Experiment 4 may be due to the addition of an enforced delay condition in Experiment 5. When dual drivers were subject to a three-second delay, they were overall faster than car drivers. This could have raised the average time taken overall by dual drivers as half of the dual drivers were subject to the three-second delay. The knock-on effect of matching accuracy on overall time taken was also discussed – dual drivers were overall faster than car drivers when subject to the three-second delay likely because dual drivers encountered the three-second delay less frequently than car drivers given their superior matching accuracy. Therefore, further evidence is needed before concluding that dual drivers also have an advantage for overall time taken to complete the motorcycle Pelmanism task. Future studies should consider the intricacies involved in a game of Pelmanism and how the number of attempts and time taken may affect each other.

Apart from attempting to replicate Experiment 4, Experiment 5 also attempted to investigate why dual drivers were not quicker at completing the motorcycle Pelmanism task. The overall time taken was broken down according to the three stages of *match selection*, *mismatch* and *match delay* during the Pelmanism game. It was hypothesised that dual drivers would take more time during the *mismatch stage* than car drivers, resulting in the observed speed-accuracy trade-off. However, Experiment 5 found no difference between driver groups in time spent on any of the stages within the Pelmanism game. Although the difference between the means did not reach statistical significance, the means can help identify a new direction for future research. For instance, looking at the means in the non-enforced condition, dual drivers consistently spent more time during the *match selection* and *match delay stage* than car drivers. Even though the difference in time spent on each stage between dual drivers and car drivers was not statistically significant, the difference could have accumulated and eventually outweighed the dual driver advantage in matching accuracy, in which shorter times are expected from fewer attempts. Therefore, it is possible that a combination of increased time spent on the *match selection* and *match delay stages* rather than the *mismatch stage* accounted for the speed-accuracy trade-off. Based on this reasoning, dual drivers may be more accurate than car drivers because they spend more time processing the first tile and thinking about its match during the *match selection stage*, as well as thinking about where to click for the next attempt during the *match delay stage*. Future studies could also examine whether the increased the time spent on the *match selection* and *match delay stages* are due to strategic differences or familiarity-based differences between driver groups.

Having established that the motorcycle Pelmanism task and its stimuli demonstrated sensitivity to players' familiarity with motorcycles, the following empirical chapter concerns the development of the motorcycle Pelmanism task as a training tool. The studies aim to investigate the optimal number of rounds that should be included in the final training tool, in order to maximise players' performance while maintaining their level of motivation. In addition, as Crundall et al.'s (2017) version of the motorcycle Pelmanism task did not include any game elements, the following studies also examine the effect of adding game elements such as goals, points and competition on players' motivation.

Chapter 6 – Development of Pelmanism

Chapter overview

In the previous chapter, the sensitivity of the motorcycle Pelmanism task was validated by assessing dual drivers' performance on the motorcycle Pelmanism task. To continue developing the motorcycle Pelmanism task as a training tool, the number of rounds to be included in the final training tool must be determined empirically. This can be achieved by presenting players with multiple rounds and identifying where players reach a performance plateau in which there is little growth or further improvement in performance. However, a potential challenge that players may lose motivation before reaching the performance plateau. So far, the design of the motorcycle Pelmanism task used in the thesis and that reported in Crundall et al.'s (2017) study consists only of the matching task as the core game mechanic and does not include any game elements. As explained in Chapter 4, game mechanics are the fundamental building blocks of the game that allows it to function. Game elements are decorative elements that enhance the gameplay experience and while they do contribute to a fully-fledged game, game elements alone do not uphold the game. Without game elements that specifically target players' motivation, drivers may quickly lose motivation to engage with the motorcycle Pelmanism-based training task, particularly when the novelty of the matching task gradually wears off. The lack of motivational features therefore poses a challenge for the real-life application of the motorcycle Pelmanism training tool. In light of this, the motorcycle Pelmanism task has scope to add different game elements in its design to enhance driver engagement with the training content and this chapter also explores the use of game elements in the motorcycle Pelmanism task.

Experiments 6 and 7 reported in this chapter examined the effects of game elements on players' motivation and performance on the matching task, with a focus on specific performance goals, rewards and competition. As highlighted in the literature review in Chapter 4, the context in which rewards and competition are used plays a crucial role in determining their effectiveness (see Cheong et al., 2014; Hanus & Fox, 2015). Therefore, it is important to examine whether the use of rewards and competition would be effective in the motorcycle Pelmanism task. Given that dual drivers consistently exhibited superiority in matching accuracy in the previous chapter, Experiment 6 examined the effectiveness of game elements that are designed to encourage accuracy. The intention was to encourage players who are not familiar with motorcycles to increase their matching accuracy during gameplay in order to attain a similar level of performance to dual drivers. However, it was found that participants were not responsive towards game elements that encouraged accuracy in the context of the Pelmanism game. To investigate whether encouraging speed would be more beneficial to players' performance and motivation, Experiment 7 modified the game elements to focus on time taken. The results showed that time-based game elements were effective in maintaining players' motivation and the plateau in players' performance was identified.

6.1 Experiment 6: Evaluation of game elements in the motorcycle Pelmanism task – encouraging accuracy

6.1.1 Introduction

The previous chapter demonstrated that the motorcycle Pelmanism task was sensitive to dual drivers' increased familiarity with motorcycles. This is promising because it indicates that the skills required by the motorcycle matching task correspond to the level of exposure to and knowledge of motorcycles in the real world. Before the effectiveness of the motorcycle Pelmanism training tool is evaluated in the final study of the thesis, research is needed to inform how many rounds of Pelmanism to include in the training tool. Given that drivers are more likely to be willing to take part in a shorter than longer training intervention, the number of rounds included in the training should be kept minimal. Previously, Crundall et al.'s (2017) version of the motorcycle Pelmanism-based training tool consisted of five rounds. However, the authors presented no empirical justification for the use of five rounds and the optimal number of rounds that should be included in a motorcycle Pelmanism-based training tool has not been tested empirically. The optimal number of rounds can be determined by when players' performance on the motorcycle Pelmanism task plateaus and motivation begins to drop significantly. This is because the plateau indicates that further engagement with the training is likely to lead to diminishing returns in performance and motivation. This rationale is similar to a previous study, which compared the change in performance across numerous training sessions (Jaeggi et al., 2008). The current design of the motorcycle Pelmanism task consists only of the matching task and because it is a simple and repetitive game, players' motivation and engagement with the matching task may diminish over time. To address this, game elements should be employed in the motorcycle Pelmanism task to enhance gameplay and sustain player's motivation before they reach a plateau in performance.

As mentioned previously, game elements contribute to players' flow state during game play, which refers to a state of optimal enjoyment and immersion in the game (Csikszentmihalyi, 1990; Nakamura & Csikszentmihalyi, 2002). Appropriate use of game elements can therefore facilitate training outcomes in the motorcycle Pelmanism task by keeping players engaged and motivated with the training content. To ensure that the game elements are effective in motivating players without compromising performance, there is a need to evaluate the impact of game elements in the context of the motorcycle Pelmanism task on players' motivation as well as performance. This can be achieved by measuring players' motivation or willingness to continue playing and the two performance measures (i.e., number of attempts and time taken to complete each round) across 10 rounds. Given that there are 12 pairs of motorcycles to match in each round, 10 rounds are likely to be sufficient to observe a plateau in performance and a significant drop in motivation. Several potential game elements have been identified during the literature review in Chapter 4, including the use of clear and specific goals, difficulty levels, competition and incentives (Bostan, & Öğüt, 2009; Fülöp, 2009; Locke & Latham, 2002; Mutter & Kundisch, 2014; Nakamura & Csikszentmihalyi, 2002). These game elements were identified because of their potential to facilitate players' flow state.

Previously, participants in Crundall et al.'s (2017) study were asked to match motorcycles (or fruits in the control condition) as quickly and as accurately as possible. This corresponds to a 'do-your-best' goal, which is not necessarily the most effective type of goal for improving motivation or performance (Locke & Latham, 2002). Without a clear target or end goal to aim towards, players can quickly lose interest and their state of flow as they do not feel appropriately challenged (Nakamura & Csikszentmihalyi, 2002). In a study comparing the effects of specific goals and do-your-best goals in a gamified task, participants who were given specific goals were more committed to attaining this goal than participants who were asked to perform as best as they could (Landers, Bauer, & Callan, 2017). Importantly, feedback on the player's progress towards the goal should also be provided so that they can reflect and adjust their efforts appropriately. With do-your-best goals, another disadvantage is that relevant feedback cannot be provided without an external frame of reference, given that such goals are often defined idiosyncratically (Locke & Latham, 2002). Not only would different players have different conceptualisations of what their best performance would be, it is also difficult to objectively measure whether a player's performance is actually his/her best. For example, one player may view do-your-best as putting in the most effort whereas another player may consider do-your-best as constantly improving his/her personal best. As a result, it is difficult to provide constructive feedback on how players can improve their performance in relation to their goal. Therefore, game elements should aim to set clear and specific goals while also providing feedback on players' progress in relation to the goal.

In serious games, the use of specific goals alone is unlikely to be sufficient to ensure self-motivated and repetitive play. Simply presenting a list of goals to attain can be uninteresting and may not be sufficient to create goal commitments. Goals should therefore be presented in an engaging way to increase players' intrinsic motivation and goal commitment. Therefore, specific goals should be embedded within other game design elements such as rewards and competition. For instance, players can be rewarded for attaining certain goals and competition can be used to introduce conflict during goal attainment. Rewards in the form of points or badges signify accomplishment of goals and can promote feelings of self-efficacy (Landers, Armstrong, & Collmus, 2017; Mutter & Kundisch, 2014), which increases the amount of effort and persistence put into attaining the goal (see Bandura, 1994). Competition allows players to demonstrate competence and according to the self-determination theory, intrinsic motivation is enhanced when players feel competent (Ryan et al., 2006). Therefore, competitive goals and goals that promise rewards can engage players and enhance their motivation.

However, there is also evidence that rewards and competition have potentially negative consequences. Rewards that are seen as controlling, such as those offered in exchange for certain behaviours, lower players' perceived autonomy as they experience pressure to act in particular ways or perform certain actions (e.g., Hanus & Fox, 2015; Mamykina et al., 2011). According to a meta-analysis by Pierce and Cameron (2002), rewards that are given for doing a task have a negative impact on self-reported interest and subsequent motivation to engage with the task because they may be seen as demands. In contrast, rewards that are given for surpassing certain standards or goals, do not undermine subsequent motivation and lead to high self-reported interest in the task because they serve as evidence of achievements. The level of autonomy associated with the rewards used in serious games is thus crucial for enhancing intrinsic motivation and should be maximised (see also Ryan & Deci, 2000). Rewards can be offered with choices, for example by allowing players to choose their reward. Therefore, to avoid offering rewards that are seen as controlling, rewards should reflect players' achievements by providing information on their performance and choices should be incorporated within the rewards. Difficulty-dependent rewards satisfy both criteria, by offering more rewards for increasing levels of difficulty and allowing players to choose which level of difficulty to aim for and the appropriate reward to receive. There is evidence that difficulty-dependent rewards can enhance motivation; participants who were offered higher rewards for increasing levels of performance subsequently spent more time on a puzzle and reported greater interest than participants who were not offered rewards (Cameron et al., 2005). Therefore, difficulty-dependent rewards are likely to be more effective than controlling rewards within the motorcycle Pelmanism-based training task.

The use of competition as a game element can also have negative consequences. Competition can be destructive in serious games when the goal is to win and beat other opponents. This is because learning may be overshadowed by the desire to outperform others in excessively competitive games that only celebrate winners. One example is the use of leader boards, which is popular in serious game designs and gamification to introduce competition (Hamari et al., 2014). Leader boards promote upward and downward comparisons because the ranking encourages players to compare themselves to players higher and lower on the board. Excessive social comparisons can however undermine players' intrinsic motivation, as evident from studies examining the use of leader boards on students' motivation on a course (Hanus & Fox, 2015; Vandercruysse et al., 2013). Furthermore, leader boards can be especially detrimental for low-performing students because social comparisons generate anxiety from being evaluated (e.g., Cheong et al., 2014). Therefore, competitive game elements should minimise the extent of social comparison during competition. According to Fülöp (2009), competitions should focus on encouraging self-improvement or mastery of goals, with little focus on other opponents in the competition. A compromise between introducing competitiveness while minimising the negative effects of social comparison is to compare players' performance to the average scores of previous players. This allows players to be competitive without focusing on winning or losing to a particular player.

Current study

The aim of the current study is two-fold. The first aim is to determine the optimal number of rounds to include in the motorcycle Pelmanism task as a training intervention for drivers. As mentioned in the introduction, this can be achieved by determining the round in which performance begins to plateau and motivation drops significantly. Measuring players' motivation to continue to the next round of Pelmanism can inform us the number of rounds before players begin to lose motivation. Performance on the motorcycle Pelmanism task is measured by number of attempts and overall time taken as with Experiments 4 and 5. It is important to also consider players' performance to ensure that the training does not become unnecessarily lengthy, such that players are benefitting or improving minimally towards the end of the training session. Players' motivation and performance across 10 rounds are therefore measured. The round in which motivation drops and performance begins to plateau can be obtained by comparing each round to the subsequent round.

The second aim of the current study is to assess the impacts of specific goals, rewards and competition on players' motivation and performance on the motorcycle Pelmanism task, measured by number of attempts and time taken. This is achieved through a control condition and two experimental conditions. The *control* condition represents do-your-best goals, as participants are tasked to match the motorcycle pairs as accurately as they can. In contrast, other participants in the two experimental conditions are presented with specific goals for the number of attempts. Participants across all conditions were encouraged to focus on matching accuracy because the previous chapter found that dual drivers consistently performed better in terms of number of attempts. In the two experimental conditions, the difficulty of the goals was varied by presenting a goal that was easy to beat and another that was harder to beat. However, the two experimental conditions differ in terms of the ways in which goals attainment is encouraged. In one experimental condition, the goals were presented with difficulty-dependent rewards, meaning that participants would receive points for beating a goal and the number of points corresponds to the difficulty of the goal. Points instead of badges were chosen to represent rewards because points provide a clear, accrual indication of a player's progress throughout the game. The higher the points, the higher the skill level and better the performance exhibited by the player. The first experimental condition is thus referred to as the *points* condition. In the other experimental condition, referred to as the *competition* condition, goals were presented competitively as the average scores of previous players. Average scores were used to encourage constructive competition and reduce excessive social comparisons as players are not able to make direct comparisons to other high-performing players. Moreover, the average scores of previous players were presented as a range (e.g., "the average number of attempts is between 50 to 87 attempts") so that the level of difficulty of goals would be comparable across the *points* and *competition* conditions. To investigate the impact of specific goals, the control condition is first compared to the mean effect of the *points* condition and the *competition* condition. This is because participants in both experimental conditions were presented with specific goals. Secondly, to examine the impact of difficulty-dependent rewards and competition, the *points* condition and *competition* condition are then compared against each other.

Based on the literature reviewed, it is hypothesised that the use of goals in both experimental conditions would enhance and maintain players' motivation over more rounds as compared to do-your-best goals in the *control* condition. Given that the goal-setting theory suggests that performance can be improved through specific goals, it is also hypothesised that players in the two experimental conditions would perform better particularly in terms of number of attempts than players in the *control* condition. Comparing the *points* and *competition* conditions, motivation may still differ because rewards and competition affect motivation through different underlying mechanisms. The use of points aims to increase self-efficacy through incentives whereas competition encourages self-improvement through challenges (Fülöp, 2009; Mutter & Kundisch, 2014). In terms of performance, it is unclear whether performance on the motorcycle Pelmanism task would be influenced by the use of rewards or competition.

6.1.2 Methods

Design

A 3x10 mixed design was used. The first between-participants variable was type of game element condition with three levels (a *control* Pelmanism task without game elements vs Pelmanism with *points* vs Pelmanism with *competition*). The game elements were designed to encourage players to prioritise accuracy during matching. The second within-participant variable was number of rounds elapsed and had 10 levels because participants took part in 10 rounds. Three dependent variables were measured and were analysed separately. They were the number of attempts and time taken to match all 12 motorcycle pairs in each round as well as participants' motivation to continue playing. Participants' motivation was measured after each round on a scale of 0-10 and reflects Garris et al.'s (2002) notion that self-motivated repetitive play is crucial for successful implementation of serious games. As there was no motivation question after the game had ended, there were only 9 rounds for the motivation analysis, resulting in a 3x9 design.

Participants

Sixty participants took part in this study, of whom 25 were male and 35 were female. Participants were recruited from the university's paid participant panel and the psychology department's credit panel. Mean age of the sample was 28.88 (S.D. = 11.05) and ranged from 18 to 66. Participants were randomly assigned to one of the three conditions – control, points or competition. Age did not differ significantly between conditions ($F (2, 57) = 0.12$, $p = .88$).

Materials and stimuli

The game was coded and run on Unity. The Pelmanism task was programmed in the same way as in Experiment 4. Nine rounds of Pelmanism were created in addition to the motorcycle round used in Experiments 4 and 5. To create nine additional sets of playing tiles, each motorcycle image was re-assigned to nine other different backgrounds such that all playing tiles had a different motorcycle-background combination. The size of the motorcycles within each pair was also varied so that participants could not match pairs based on backgrounds and sizes. The height of the motorcycles ranged between 17%-50% of the height of the image and were located between the foreground or middle ground of the picture. All participants saw the same set of playing tiles in each round. The order of the rounds and their respective set of playing tiles were presented in the same order but the positions of playing tiles in each round were randomised across all participants. Several changes were made to the game in the *points* condition and *competition* condition, namely the addition of colour to the attempts counter to reflect the targets presented. This is further explained in the respective sections below.

Control condition

The *control* condition is similar to the version of the Pelmanism task used in Experiments 4 and 5 because no game elements were used. Nevertheless, the matching task was described as a game to participants to account for potential differences in perception of the game between conditions. This is because players' experience of an activity could be affected by whether the activity is framed as a game or task (Lieberoth, 2015). Participants were instructed to match the pairs as accurately as they could. The colour of the stopwatch timer and attempts counter were black throughout the game, similar to Experiments 4 and 5. Participants' time taken and attempts made were shown after each round and they were asked to rate how motivated they were to continue to the next round using the keyboard. At the end of 10 rounds, a summary of their performance was shown on the screen.

Points condition

The two experimental conditions, namely Pelmanism with *points* and Pelmanism with *competition*, used points and competition respectively to encourage participants to match motorcycle pairs as accurately as they could. Participants in the *points* condition scored points based on the goals for the number of attempts to complete each round. At the beginning of each round, participants were shown two goals – a hard and an easy goal. The goals were the number of attempts to beat in each round and would become increasingly difficult as the rounds continued. Participants would earn 10 points for beating the hard goal or 5 points for beating the easy goal but not the hard goal. If they completed the round without beating the easy goal, they would receive only 2 points. The goals were determined based on data from three pilot participants (mean age = 44.6 years, 2 male), who each completed ten rounds of the motorcycle Pelmanism task as described in the control condition. A regression line was fitted to the data for mean number of attempts, in order to derive the hard and easy goals for each round (refer to Appendix F). The goals therefore became increasingly challenging with subsequent rounds. The final goals are shown in Table 6.1.

Table 6.1: Hard and easy goals to beat in each round

Round	Hard goal	Easy goal
1	50	87
2	49	83
3	48	79
4	47	75
5	46	71
6	45	67
7	44	63

8	43	59
9	42	55
10	41	51

To provide feedback during the game, the counter changed colour based on participants' performance in relation to the goals. For instance, in Round 1, the counter would be green as long as participants made fewer than 50 attempts (hard goals) but would change to yellow if participants made more than 50 attempts. If they made more than 87 attempts (easy goals), the counter changed to red. Their number of attempts, time taken and number of points earned were shown at the end of each round as well as the end of the game.

Competition condition

Competition was introduced by presenting the average number of attempts required by previous players and encouraging players to beat the scores of previous players. To ensure the same level of difficulty as the *points* condition, the hard and easy goals became the lower and upper boundaries of the average number of attempts required by previous players (see Table 6.1 above). This means that participants were shown the same goals across the two experimental conditions. However, instead of earning points for beating a goal, participants were given feedback on whether they outperformed the average scores of previous players. For example, in the first round of the *competition* condition, participants were shown that the average score of previous players was between 50 to 87 attempts. Participants perform better than the average by making fewer than 50 attempts or worse than average if they made more than 87 attempts. If they made between 50 to 87 attempts, they were on par with the average scores. Similar to the *points* condition, there was also a colour-changing counter during the game. The counter would be green as long as participants were performing better than average but would change to yellow if participants were performing on par with the average scores. If they could not beat the upper boundary of the average scores, the counter changed to red. Their number of attempts, time taken and performance relative to the average were shown at the end of each round as well as the end of the game.

Procedure

Participants were randomly allocated to one of the three conditions. A pictorial example of a matched pair was given at the start to emphasize that participants could not use the background or the size of the motorcycle as heuristics to help them match the pairs. The example matched pair was of cars rather than motorcycles, to avoid any priming effects of certain motorcycle models. The appropriate instructions for each condition were given both on the screen and verbally. Participants were informed that there would be a total of 10 rounds and that the timer would only start on the first click in each round. They also had the opportunity to clarify any questions about the game before starting. At the end of each round, participants were asked how motivated they were to continue to the next round and to rate their answer on a scale of 0-10 using the keyboard. Participants received feedback in terms of their scores, along with the points or performance in relation to average where applicable, at the end of each round and at the end of the game (see Figure 6.1 for an example of a summary of participants' performance at the end of the game).

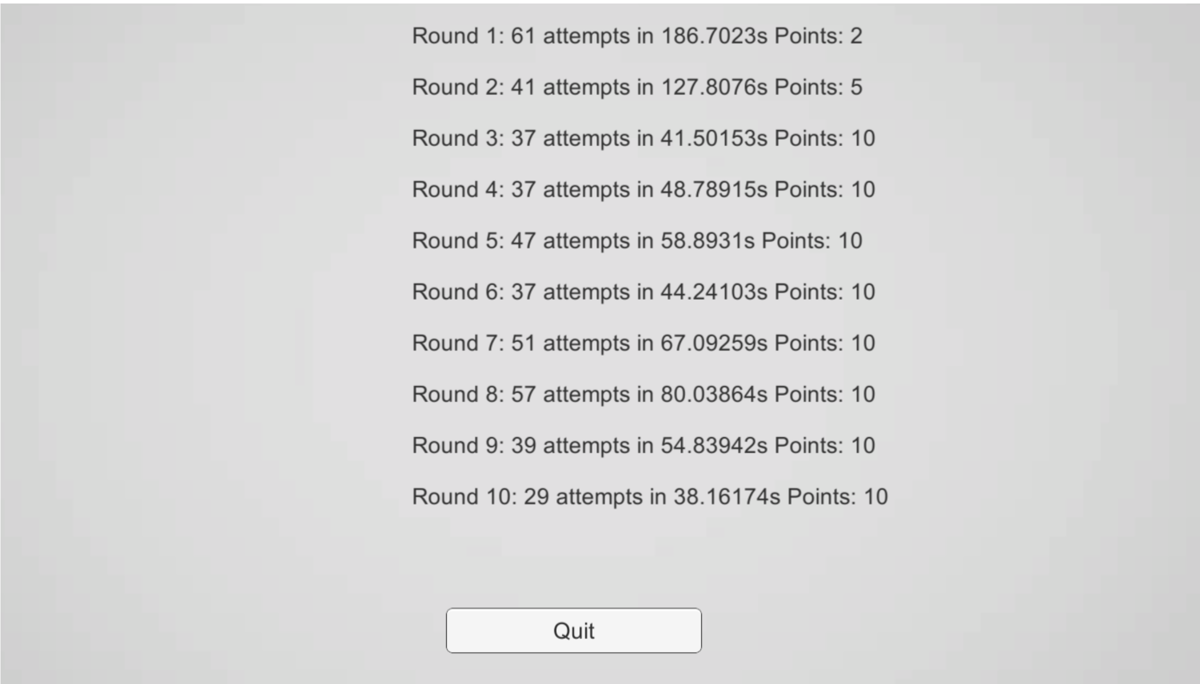

Figure 6.1: Example of feedback received by participant at the end of 10 rounds, and in this case, participant was in the points condition.

Statistical analysis

Separate analyses were conducted for each of the three dependent variables – motivation, number of attempts and time taken. Data were analysed using the statistical programming language R (R Core Team, 2020). Number of attempts and time taken were analysed using linear mixed effects models (refer to Chapter 2). However, as motivation is an ordinal variable, a cumulative link mixed model (clmm) from the ordinal package (Christensen, 2019) was used. When the data are categorical and the categories follow a clear order, ordinal regression should be used. Linear regression should not be used because the dependent variable is no longer continuous or normally distributed, which violates the assumptions of linear models (Bürkner & Vuorre, 2019). The cumulative link from the clmm takes into account the order of the response categories, similar to how a binomial link is used to reflect a dichotomous dependent variable. The method of constructing clmm models and analysing fixed effects using likelihood ratio tests is the same as linear mixed effects models. The full model for each dependent variable with by-participant random effect was:

Motivation/ Attempts/ Time taken ~ (1|Participant) + Round + Condition + Round:Condition

The full model for each dependent variable performed better than the respective general linear model without random effects (refer to Appendix G).

In contrast to the default dummy coding scheme (see Chapter 2), a successive difference coding scheme was used for the within-participant variable of rounds elapsed in order to compare each round with the subsequent round. This resulted in 9 contrasts (8 for motivation). A helmert coding scheme was used for the between-participant variable of game element condition to conduct two contrasts – 1) control condition vs mean effect of *points* and *competition* conditions, and 2) *points* condition vs *competition* condition. This allows us to first compare the effectiveness of specific goals in the experimental conditions and do-your-best-goals in the *control* condition, and then compare the effectiveness of points and competition as incentives for goal attainment. The planned contrasts were set up to be orthogonal and thus multiple comparison corrections were not needed because each contrast is independent of each other (Abdi & Williams, 2010; Bechhofer & Dunnett, 1982).

6.1.3 Results

Motivation

The mean motivation rating for each round according to condition is shown in Figure 6.2.

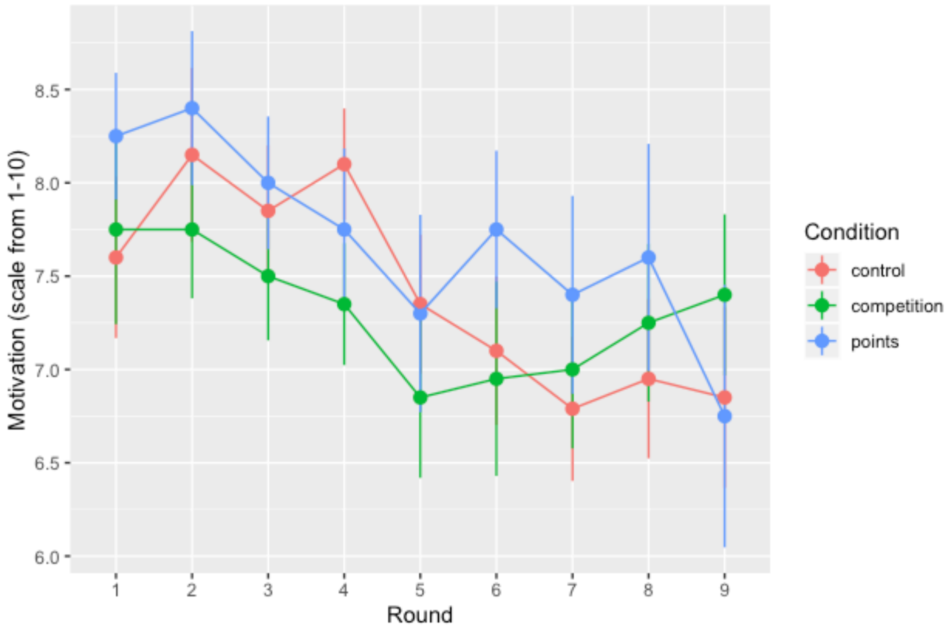

Figure 6.2: Trend in mean motivation rating in each round, rated from 1-10, by condition with +/-1 SE bars

There was a significant main effect of round in participants' motivation (χ^2 (8) = 42.58, p < .001), and the planned contrasts of successive difference revealed that the mean motivation decreased significantly from Round 4 (M = 7.73) to Round 5 (M = 7.17; z = -2.32, p = .020; β = -0.77 [95% CI: -1.42, -0.12]). The effect of condition was not significant (χ^2 (2) = 1.40, p = .50). The planned contrast between control and experimental condition was not significant (z = -0.39, p = .70; β = -0.09 [95% CI: -0.55, 0.37]), nor was the planned contrast between points and competition (z = -1.22, p = .26; β = -0.46 [95% CI: -1.25, 0.34]). The interaction between condition and round was also not significant (χ^2 (16) = 16.3, p = .43).

Number of attempts

The mean number of attempts from the first to the tenth round according to condition is shown in Figure 6.3.

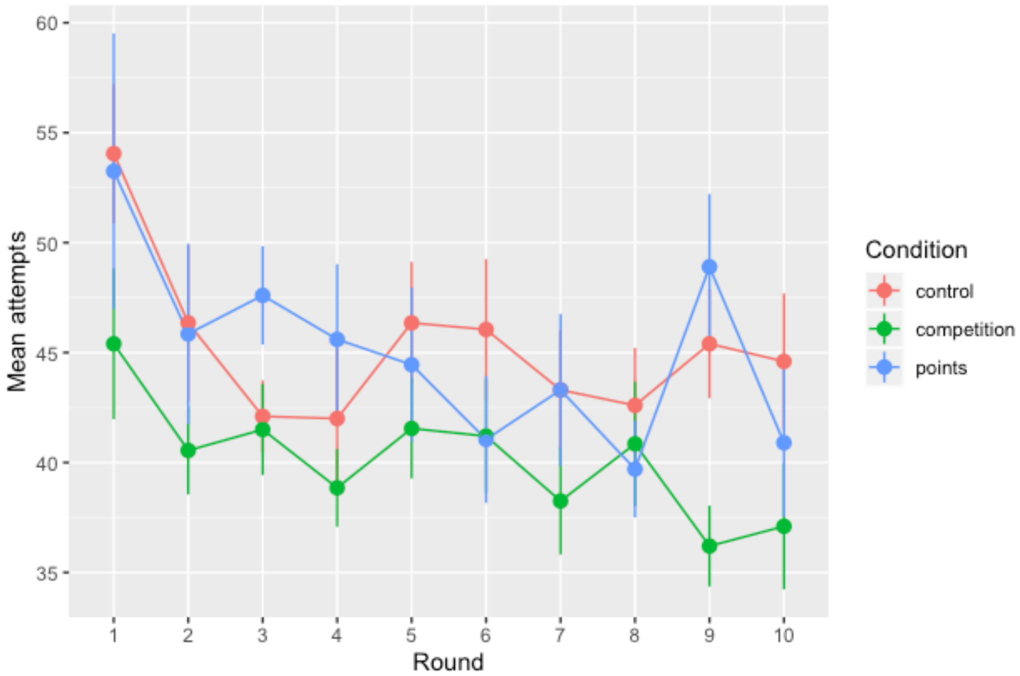

Figure 6.3: Trend in mean number of attempts in each round by condition with +/-1 SE bars

The main effect of round was significant (χ^2 (9) = 40.64, p < .001) and the planned contrasts of successive difference found that the number of attempts decreased significantly from the first round (M = 50.9, SD = 20.2) to the second round (M = 44.2, SD = 14.9; t (531) = -3.51, p < .001; β = -6.65 [95% CI: -10.4, -2.93]). There was no difference between conditions in the number of attempts averaged over 10 rounds (χ^2 (2) = 3.92, p = .14). The planned contrast between control and experimental conditions was not significant (t (57) = 1.05, p = .30; β = 0.89 [95% CI: -0.78, 2.57]), nor was the planned contrast between points and competition (t (57) = -1.66, p = .10; β = -2.46 [95% CI: -5.36, 0.44]). There was also no significant interaction between condition and round for number of attempts (χ^2 (18) = 20.48, p = .31).

In the *points* condition, the proportion of participants who beat the hard goal was 60% in the first round but increased to 75% in the last round. In the *competition* condition, the proportion of participants who beat the hard goal was 60% in the first round but increased to 95% in the last round. In the control condition, 40% of the participants would in theory have beaten the hard goal in the first round, increasing to 70% in the last round.

Time taken

The mean time taken from the first to the tenth round according to condition is shown in Figure 6.4.

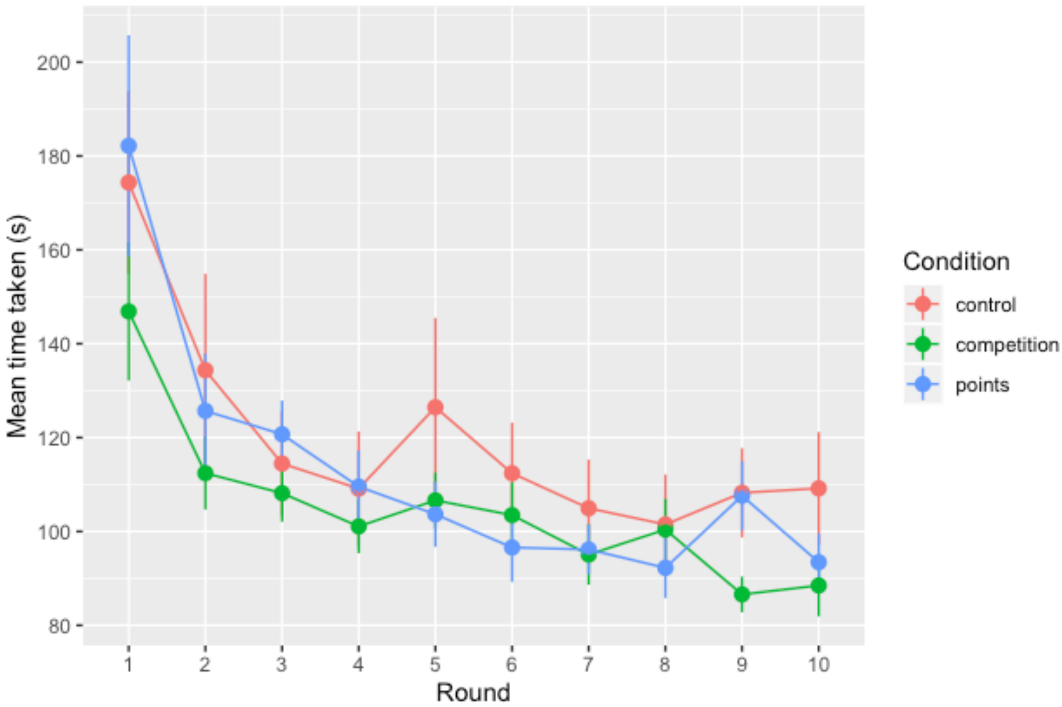

Figure 6.4: Trend in mean time taken in each round by condition with +/-1 SE bars

The assumption of normality of the residuals was violated in this model and a logarithm transformation was performed on the time taken. There was a main effect of round on the time taken to match all 12 motorcycle pairs (χ^2 (9) = 177.25, p < .001). Time taken reduced significantly from Round 1 (M = 168.0s, SD = 87.4) to Round 2 (M = 124.0s, SD = 64.6; t (531) = -6.77, p < .001; β = -0.30 [95% CI: -0.39, -0.21]). There was no difference between conditions in time taken averaged over 10 rounds (χ^2 (2) = 0.83, p = .66). The planned contrast between control and experimental conditions was not significant (t (57) = 0.65, p = .51; β = 0.016 [95% CI: -0.031, 0.063]), nor was the planned contrast between points and competition ($t(57) = -0.60$, $p = .55$; $\beta = -0.025$ [95% CI: -0.11, 0.057]). There was no significant interaction between condition and round ($\chi^2(18) = 20.7$, $p = .29$).

6.1.4 Discussion

The present study examined the effects of specific goals, points and competition in the context of the motorcycle Pelmanism task on players' motivation and performance in terms of number of attempts and time taken. Across all conditions, participants' motivation to continue playing dropped significantly from the fourth to the fifth round. This suggests that participants began to lose interest and engagement with the training content after four rounds. However, the presence of game elements in the two experimental conditions did not improve motivation and did not sustain motivation across more rounds than the *control* condition. Therefore, the hypotheses that the use of goals, points or competition would influence motivation were not supported. Given that the design of the game elements was closely informed by motivational theories and literature, this finding was unexpected. The findings demonstrate the importance of evaluating the design and implementation game elements in the context to be used rather than mimicking existing game designs because the effect of game elements can vary widely according to contexts (Seaborn & Fels, 2015). Based on the implementation in the current study, neither goals, points nor competition were effective in motivating players.

The performance measures indicated that participants across all conditions improved significantly from the first to the second round. This could suggest that a point of diminishing return in motorcycle perceptual learning is reached after two rounds, after which the amount of effort put into the training outweighs the benefits gained from the training. However, participants' performance was not affected by the presence of game elements given that there was no significant interaction between condition and round for either performance measures. Instead, it is likely that participants may have used the first round to test out different strategies because they were still unsure about the game. The improvement from the first to second round could therefore be due to procedural learning or practice effects as participants became used to the task. In light of this, we cannot be certain whether participants experienced procedural learning or an actual learning curve for motorcycle perceptual learning.

The current findings indicate that the presence of game elements did not enhance motivation and improvements in participants' performance began to level out after the second round. The game elements were therefore not effective enough to motivate players and enhance their experience of gameplay above that of the basic matching task (i.e., *control* condition). Consequently, it is difficult to ascertain whether the lack of motivating element contributed to the early performance plateau observed after the second round. This is because drivers may put less effort into the matching task if they are not motivated to engage with the training, resulting in a shallow learning curve as performance improves slowly. Without motivating game elements, drivers are unlikely to take part and voluntarily engage with the motorcycle Pelmanism-based training task or may lose interest before completing the training. This is a key challenge for the application of the motorcycle Pelmanism training intervention and other hazard perception training interventions in real life. Therefore, it is important that the game elements used are able to influence motivation accordingly and sustain players' motivation until a performance plateau is reached.

One possible reason why the game elements in the current study were not effective could be that they were designed to direct players' attention towards the number of attempts made during each round. Game elements that encourage matching accuracy may have affected motivation only when players are performing an attempt, that is between the first and second click of each attempt. For instance, players may feel motivated to recall which tile correctly matches the single tile that is currently face up and once they have clicked on the second tile, their motivation increases or decreases depending on whether the tiles match or not. However, once the second tile is turned over, participants cannot actively advance their progress to achieve the goal based on attempts, resulting in a lull in motivation. Therefore, the current design of the goals leads to small, interrupted waves of motivation that occur only when players perform a match, thus limiting the effectiveness of the game elements. An improvement to the design of the game elements could be to direct players' attention towards time taken rather than number of attempts. By presenting goals as time taken and using points or competition to encourage achievement of time-based goals, a sense of urgency can be created and capitalised on to provide a constant source of motivation for players. In addition, time-based goals are more intuitive and easier to keep track of than the number of attempts. For example, players likely have to refer to the attempts counter more frequently to keep track of their accuracy than if they were to refer to the timer to keep track of their speed. Therefore, another study should be conducted to investigate the impact of time-based goals with points or competition on motivation and performance.

In conclusion, the current study found that regardless of the presence of game elements, motivation declined significantly after the fourth round and both performance measures began to plateau after the second round. The use of attempts-based goals, points or competition did not prolong participants' motivation during the motorcycle Pelmanism task and did not impact the number of attempts nor time taken to complete each round of Pelmanism. The lack of motivational effects from the game elements used in the current study poses a challenge to the development of the motorcycle Pelmanism task as a training tool because participants could lose interest and drop out before completing the training session. It was proposed that time-based goals would be more effective than attempts-based goals because time-based goals may create a sense of urgency and thus motivate players effectively. Therefore, the next study examines whether time-based goals alongside points or competition would be effective in enhancing participants' motivation as compared to a control condition.

6.2 Experiment 7: Evaluation of game elements in the motorcycle Pelmanism task – encouraging speed

6.2.1 Introduction

In the previous study, attempts-based goals, points and competition were found to have little effect on participants' motivation and performance across 10 rounds of the motorcycle

Pelmanism task. As a result, the study could not inform the number of rounds to include in the final training tool because the lack of motivational effects from the game elements could have contributed to the early performance plateau observed after the second round. A possible reason for the lack of motivational effects eluded to the possibility that participants may have experienced small bursts of motivation only when performing a match because goals framed as number of attempts were relevant only when participants click on the second tile. Once the second tile is clicked on, participants cannot make any progress in relation to the goal until the next attempt. Furthermore, attempts-based goals are less intuitive and more effortful to adhere to than time-based goals. Therefore, the goals presented in the motorcycle Pelmanism task could be modified to reflect time taken, alongside difficulty-dependent points or competition to encourage players to finish each round as quickly as possible.

Current study

Similar to the previous experiment, the aim of the current study is two-fold. First, to determine the optimal number of rounds to include in the training and secondly, to investigate the effect of time-based goals, points and competition on players' motivation and performance over 10 rounds. Using the same methodology and experimental design, the round in which performance begins to plateau and motivation drops significantly is determined by comparing successive rounds. Three versions of Pelmanism are also used, a *control* condition, a *points* condition and a *competition* condition. Both the *points* and *competition* conditions used time-based goals but differed in terms of how goal attainment is encouraged – through difficulty-dependent points or competition respectively. Therefore, the effect of time-based goals can be examined by comparing a control do-your-best condition against the mean of two experimental conditions. The two experimental conditions are then compared against each other to investigate the effectiveness of points and competition. Based on the literature reviewed previously, it is hypothesised that the presence of game elements in the two experimental conditions would increase motivation as compared to the *control* condition. Comparing the *points* condition and *competition* condition, motivation may differ given that rewards increase self-efficacy while competition provides challenges. In terms of performance, it is unclear whether performance on the motorcycle Pelmanism task, measured by time taken and number of attempts, would be influenced by the use of time-based goals, points or competition.

6.2.2 Methodology

Design

The same 3x10 mixed design was used as in the previous experiment. There were two independent variables – number of rounds elapsed and game element (*control* vs Pelmanism with *points* vs Pelmanism with *competition*). Instead of attempts-based goals, the two experimental conditions used time-based goals. The three dependent variables were time taken, number of attempts to complete each round and participants' motivation to continue playing. Motivation was measured on a scale of 0-10 and included only 9 rounds because no motivation question was asked on the last round, thus giving a 3x9 design.

Participants

Seventy-six participants took part in this study, of whom 20 were male and 56 were female. Participants were recruited from the university's paid participant panel and the psychology department's credit panel. Mean age of the sample was 27.69 (S.D. = 11.03) and ranged from 18 to 63. Participants were randomly assigned to one of the three conditions – control, points or competition. Age did not differ significantly between conditions ($F(2, 69) = 1.51$, $p = .23$). Four participants did not provide their age.

Materials and stimuli

The same materials and stimuli were used as in Experiment 6. The setup of each condition was also the same, except for the use of time-based goals in the *points* condition and *competition* condition. Participants in the *control* condition were asked to match the pairs as quickly as they could while participants in the experimental conditions were given specific goals presented in terms of time taken (see Table 6.2). Similar to Experiment 6, the hard and easy goals were determined by calculating two regression lines using the time taken by the three pilot participants (see Appendix F). The colour of the timer changed according to the participant's performance in relation to the goals, similar to Experiment 6.

Table 6.2: Hard and easy goals to beat in each round

Round	Hard goal (min)	Easy goal (min)
1	2:08	2:40
2	2:03	2:35
3	1:58	2:30
4	1:53	2:25
5	1:48	2:20
6	1:43	2:15
7	1:38	2:10
8	1:33	2:05
9	1:28	2:00
10	1:23	1:55

Procedure

The procedure was the same as in Experiment 6, except that participants were instructed to complete each round as quickly as possible rather than as accurately as possible.

Analysis

Data were analysed in the same way as Experiment 6 – motivation was analysed using cumulative link mixed models whereas time taken and number of attempts were analysed using linear mixed models. Planned contrasts were coded in the same way and the full model is as follows:

Motivation/ Time taken/ Attempts ~ (1|Participant) + Round + Condition + Round:Condition

The full model for each of the dependent variables performed better than the respective general linear model without random effects (refer to Appendix H).

6.2.3 Results

Motivation

The mean motivation rating for each round according to condition is shown in Figure 6.5.

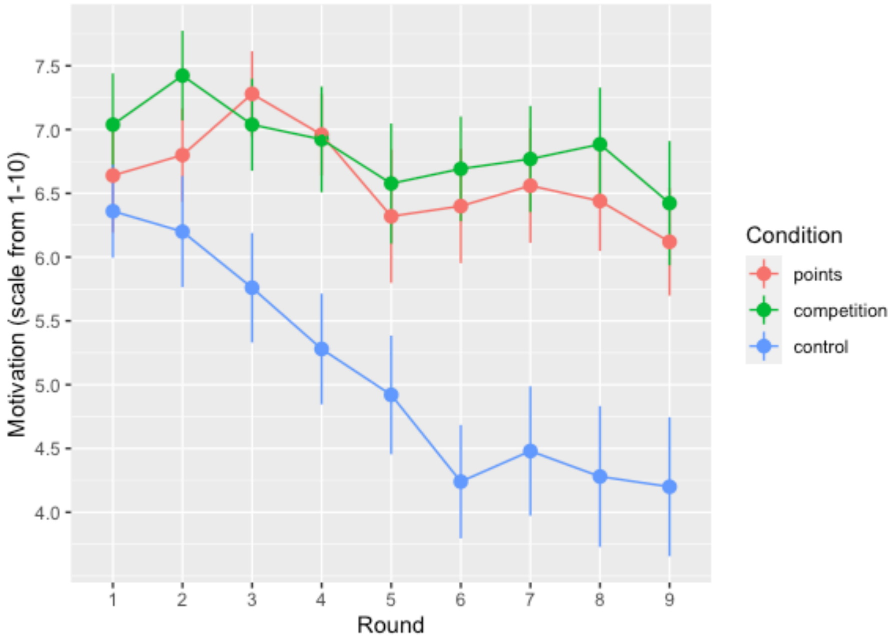

Figure 6.5: Mean motivation rating in each round, rated from 1-10, by condition with +/-1 SE bars

Likelihood ratio test found a significant main effect of round in participants' motivation (χ^2 (8) = 60.09, p < .001); mean motivation decreased significantly from Round 4 (M = 6.39) to Round 5 (M = 5.95; z = -2.00, p = .046; β = -0.55 [95% CI: -1.09, -0.010]). The effect of condition was significant (χ^2 (2) = 13.37, p = .0012). The planned contrast between control and experimental conditions was significant (z = -3.85, p < .001; β = -0.78 [95% CI: -1.17, -0.38]), with higher motivation ratings in both points (M = 6.61, SD = 2.07) and competition conditions (M = 6.86, SD = 2.12) than the control condition (M = 5.08, SD = 2.44). The planned contrast between points and competition was not significant (z = 0.92, p = .36; β = 0.30 [95% CI: -0.34, 0.93]). The interaction between condition and round was significant (χ^2 (16) = 32.67, p = .0082). Motivation in the control condition was significantly lower than the experimental conditions from the third through to the last round (see Figure 6.5 for means and Table 6.3 for contrasts).

Table 6.3: Significant difference between control and experimental conditions from Round 3 - Round 9 (below significance threshold of .05)*

Round	z-value	β [95% CI]	*p-value*
3	-2.18	-0.48 [-0.92, -0.048]	.029*
4	-2.67	-0.59 [-1.02, -0.16]	.007*
5	-2.57	-0.57 [-1.0, -0.13]	.010*
6	-3.79	-0.83 [-1.26, 0.40]	< .001*
7	-3.47	-0.78 [-1.23, -0.34]	< .001*
8	-3.51	-0.80 [-1.25, -0.35]	< .001*
9	-3.01	-0.69 [-1.14, -0.24]	.0026*

Time taken

The mean time taken to complete each round according to condition is shown in Figure 6.6.

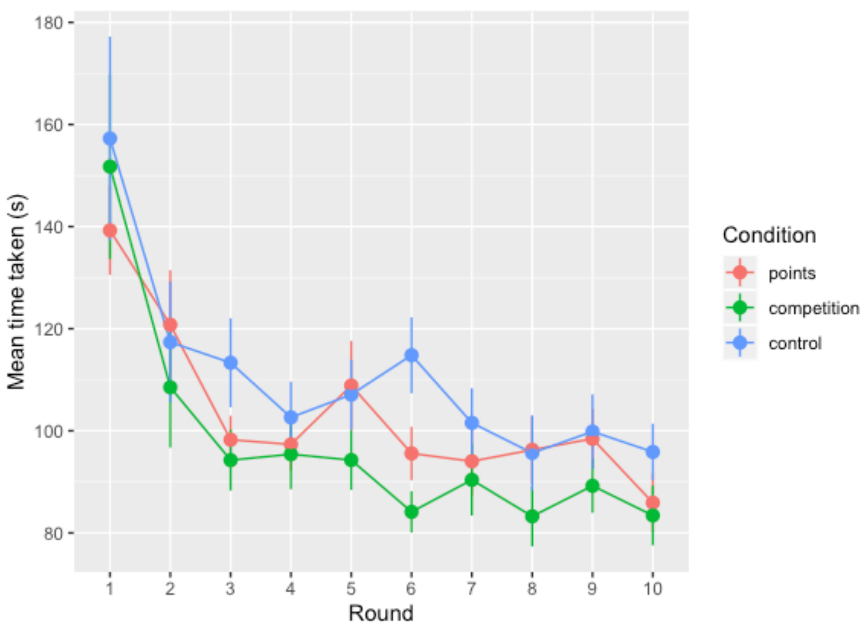

Figure 6.6: Trend in mean time taken in each round by condition with +/-1 SE bars

A logarithm transformation was performed on time taken because assumption of normality of the residuals was violated. There was a main effect of round in the time taken to match all 12 motorcycle pairs (χ^2 (9) = 183.18, p < .001). Time taken reduced significantly from Round 1

(M = 149.0s, SD = 81.8) to Round 2 (M = 115.0s, SD = 57.2; t (675) = -6.39, $p < .001$; β = -0.25 [95% CI: -0.33, -0.18]) and then from Round 2 to Round 3 (M = 102.0s, SD = 34.0; t (675) = -2.07, $p = .038$; β = -0.082 [95% CI: -0.16, -0.0045]). Time taken decreased again from Round 9 (M = 95.8s, SD = 31.0) to Round 10 (M = 88.3s, SD = 29.0; t (675) = -2.09, $p = .037$; β = -0.083 [95% CI: -0.19, -0.0053]). The main effect of condition was not significant (χ^2 (18) = 3.59, $p = .16$). The planned contrast between control and experimental conditions was non-significant (t (73) = 1.47, $p = .15$; β = 0.030 [95% CI: -0.010, 0.071]), nor was the planned contrast between points and competition (t (73) = -1.16, $p = .25$; β = -0.041 [95% CI: -0.11, 0.029]). There was no significant interaction between condition and round (χ^2 (18) = 14.89, $p = .67$).

In the *points* condition, the proportion of participants who beat the hard goal was 52% in the first round but increased to 88% in the last round. In the *competition* condition, the proportion of participants who beat the hard goal was 61% in the first round but increased to 92% in the last round. In the control condition, 44% of the participants would in theory have beaten the hard goal in the first round, increasing to 84% in the last round.

Number of attempts

The mean number of attempts to complete each round according to condition is shown in Figure 6.7.

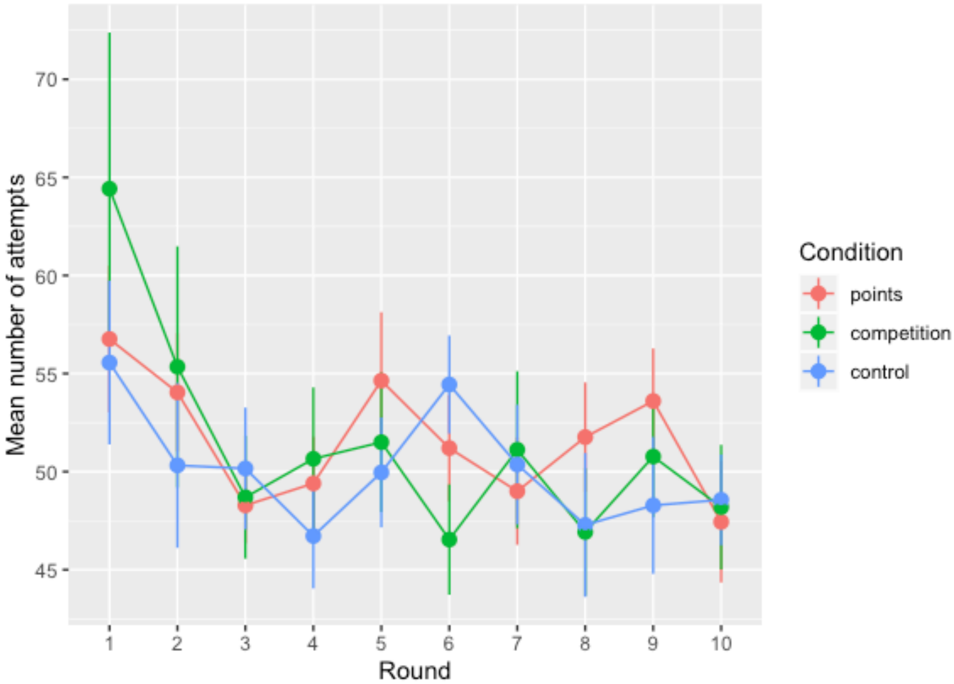

Figure 6.7: Trend in mean number of attempts in each round by condition with +/-1 SE bars

The main effect of round was significant for the number of attempts needed to match all motorcycle pairs (χ^2 (9) = 38.49, $p < .001$). The number of attempts decreased significantly from the first round (M = 59.0, SD = 28.6) to the second round (M = 53.3, SD = 23.4; t (675) = -3.51, p = .0089; β = -5.72 [95% CI: -10.0, -1.45]). There was no difference between conditions in the number of attempts averaged over 10 rounds (χ^2 (2) = 0.20, p = .91). The planned contrast between control and experimental conditions was non-significant (t (73) = -0.43, p = .67; β = -0.45 [95% CI: -2.49, 1.59]), nor was the planned contrast between points and competition (t (73) = -0.055, p = .96; β = -0.098 [95% CI: -3.61, 3.41]). There was no significant interaction between condition and round for number of attempts (χ^2 (18) = 19.09, p = .39).

6.2.4 Discussion

The present study examined the effects of time-based goals, points and competition on players' motivation and performance across 10 rounds. The results found that participants in the *points* and *competition* conditions were significantly more motivated than participants in the *control* condition. The hypothesis that motivation would be enhanced by the presence of game elements in the experimental conditions is thus supported. Furthermore, the game elements were able to sustain participants' motivation across more rounds than in the *control* condition. Participants in the *points* and *competition* conditions were significantly more motivated than participants in the *control* condition from the third round onwards, during which a significant drop in motivation in the *control* condition was observed. The findings indicate that specific goals, alongside rewards or competition, are more effective in enhancing motivation than do-your-best goals. This is in line with previous evidence demonstrating an increase in commitment and effort towards attaining specific goals as compared to do-your-best goals (Landers & Landers, 2014; Landers, Bauer, & Callan, 2017). Comparison of participants' motivation ratings in the *points* condition and *competition* condition found little difference. However, given that motivation in either conditions was still higher than the *control* condition, this suggests that both difficulty-dependent points and competition are suitable game elements to encourage goal attainment in the motorcycle Pelmanism task.

In terms of performance, the number of attempts significantly improved from the first to the second round, which is likely to be due to procedural learning as observed in the previous experiment. Time taken improved significantly from the first to the third round across all conditions and again from the ninth to the tenth round. It is, however, possible that participants performed better in the tenth round because they knew it was the final round and were therefore motivated to finish it as quickly as possible. Nevertheless, the improvements across the first three rounds indicate that participants experienced the steepest learning curve during the initial three rounds. Together with the finding that the impact of game elements on motivation is evident from the third round, three rounds are potentially an appropriate number of rounds to include in the final training tool. This is because the game elements can help mitigate the drop in motivation while ensuring that participants experience a steep learning curve. The advantage of a short training session is that drivers would be willing to participate in the training, particularly as driver training interventions are usually non-compulsory.

A strength of this study is the careful design and implementation of game elements that were grounded in theory. Specifically, the use of difficulty-dependent points allows the training tool to focus on rewarding players appropriately to increase self-efficacy and facilitate perceived autonomy whereas the use of competition allows the training tool to encourage self-improvement without excessive social comparison (Fülöp, 2009; Ryan & Deci, 2000). In cases where game elements were applied inappropriately, there can be little motivational benefits and could even led to undesirable consequences. For instance, Mamykina et al. (2011) found that users on StackOverflow ceased to participate after reaching 1000 points on the platform because no additional benefits were offered. In another study investigating the use of points and leader boards on an image tagging platform, no effect on participants' motivation was found, likely because points were offered in exchange for tagging behaviour rather than the quality of their tags (Mekler et al., 2017). Furthermore, the leader board used in their study displayed the rankings of all players and encouraged excessive comparisons with other players. These limitations were overcome in the current study by adapting the points scoring system to reflect the difficulty of the goals and balancing the level of competitiveness. Landers, Armstrong, & Collmus (2017) also noted that the uptake of game elements should feel optional to players, otherwise they quickly become a requirement of the learning process, which could undermine players' motivation to engage with the learning content. The current study was able consider players' perceived autonomy as participants were presented with the freedom to choose which difficulty goal to aim for.

To conclude, the motivational benefits of game elements were demonstrated in the present study. Compared to a *control* condition without game elements, introducing competition and rewarding players with points to attain time-based goals led to higher motivation across more rounds. In the third round, participants lost their motivation to continue playing when game elements were not present, as demonstrated by the *control* condition, but this was ameliorated by the use of time-based goals, points and competition in the two experimental conditions. The third round was also when the time taken by participants to complete the round began to plateau and no substantial improvements in participants' performance were observed in subsequent rounds. Therefore, it was proposed that three rounds would be an appropriate number of rounds in the motorcycle Pelmanism training tool, and this will be evaluated in the next chapter. There was no difference in motivation between the use of points and competition, which suggests that participants were equally motivated when rewards were offered or when given the opportunity to compete with others during the motorcycle Pelmanism game.

6.3 General discussion

This chapter reported two experiments conducted to evaluate the effects of game elements, namely specific goals, points and competition, in the context of the motorcycle Pelmanism task. In the first experiment, the design of the game elements encouraged players to match motorcycle pairs from memory as accurately as possible. The findings suggest that this design had little influence on players' motivation and performance, particularly in terms of the number of attempts made. The second experiment modified the design of the game elements such that they encouraged players to match motorcycle pairs from memory as quickly as possible. The modification resulted in higher and prolonged motivation reported by players in the experimental conditions than the *control* condition and a steeper learning curve for time taken in the initial rounds compared to the previous experiment. Based on the current findings, the most immediate recommendation for the development of a motorcycle Pelmanism training tool is to use game elements such as time-based goals, difficulty-dependent points and/or competition to enhance drivers' motivation to engage with the training. This is because attempts-based goals did not impact motivation nor performance whereas time-based goals did. Furthermore, difficulty-dependent points and competition were similarly effective in terms of encouraging players to achieve the goals presented during the Pelmanism-based training task.

Another significant contribution of the findings pertains to the number of rounds that the motorcycle Pelmanism training task in the final study should include. As mentioned in the introduction, the number of rounds to be included should be determined by the point at which participants' motivation drops and performance begins to plateau. The benefits of the game elements were evident from the third round onwards as motivation was significantly higher in both experimental conditions than in the *control* condition. This indicates that the game elements helped sustain participants motivation, which would have otherwise declined significantly from the third round. Using time-based goals, the time taken by participants to complete each round also began to plateau after the third round. Although there is an argument for maximising the benefits of game elements by including all 10 rounds in the final study, there is no need to do so because participants' performance generally did not improve any further after the third round. Therefore, a sensible number of rounds to include in the final evaluation of the motorcycle Pelmanism training task should be three. This deviates from the design of Crundall et al.'s (2017) motorcycle Pelmanism task as they included five rounds rather than three. However, a shorter training session is likely to be more appealing to drivers than a longer training session, which would increase the likelihood that drivers choose to take part in the training intervention. Otherwise, the training intervention may have little real-world impact even if it demonstrates strong training effects in research settings. Therefore, evaluating the effectiveness of a shorter motorcycle Pelmanism training task has real-world implications.

Having empirically examined the number of rounds and the design of game elements, the next chapter assesses the effectiveness of the motorcycle Pelmanism training task shaped by the findings of this chapter. A range of the perceptual tasks including the hazard prediction task, visual search task and T-junction task will be used to evaluate the effectiveness of the motorcycle Pelmanism training task in improving drivers' ability to detect and identify motorcycles.

Chapter 7 – Evaluation of motorcycle Pelmanism task as a training tool

Chapter overview

In this final empirical chapter, an improved version of Crundall et al.'s (2017) motorcycle Pelmanism task as a training tool is evaluated. This version was based on previous findings in this thesis. In Chapter 5, the game mechanics and stimuli used in the motorcycle Pelmanism task demonstrated sensitivity to motorcycle familiarity across two experiments, indicating that the training task taps into relevant skills that reflect the level of exposure to motorcycles in real life. In Chapter 6, the effects of goals, points and competition as game elements in the motorcycle Pelmanism task were investigated. The appropriate use of points and competitions to encourage time-based goals effectively maintained players' motivation without compromising time taken and number of attempts. To evaluate the effectiveness of the motorcycle Pelmanism task as a training tool, a range of perceptual tasks taken from Chapter 3 was used to measure drivers' motorcycle perceptual skills before and after completing the motorcycle Pelmanism training task.

7.1 Experiment 8: Effectiveness of the motorcycle Pelmanism task as a training tool

7.1.1 Introduction

In an innovative application of the Pelmanism game to perceptual training, Crundall et al. (2017) reported that a motorcycle Pelmanism-based training task was sufficient to improve drivers' ability to detect motorcycles in a T-junction task. To further investigate and develop the motorcycle Pelmanism task as a training tool, an in-depth examination of the motorcycle Pelmanism-based task was conducted in this thesis. A series of experiments was conducted in Chapters 5 and 6 to examine the sensitivity of the game mechanic of the Pelmanism, the number of rounds to include, as well as the effectiveness of game elements in enhancing the players' motivation during the motorcycle Pelmanism training task. The motorcycle Pelmanism task demonstrated sensitivity to players' familiarity with motorcycles. Dual drivers, who had more exposure to and were more familiar with motorcycles, consistently performed better than car drivers in terms of the number of attempts needed to match all motorcycle pairs. This indicated that the motorcycle matching task relies on skills pertinent to real world motorcycle familiarity or experience. The ecological validity of the motorcycle Pelmanism task was also improved relative to Crundall et al.'s (2017) version as motorcycles with riders embedded within road backgrounds were used as matching pairs. This was more naturalistic than Crundall et al.'s (2017) version, which used motorcycles without riders on a white background as matching pairs. The authors had previously pointed out that the motorcycle Pelmanism task has scope to include game elements to increase motivation and engagement during training, but no game elements were used in their study. A literature review conducted on serious games in Chapter 4 helped identify potential game elements that contribute to the flow state of players. Chapter 6 then investigated the effects of game elements such as goals, points and competition on players' motivation and performance on the motorcycle Pelmanism game. The results found that the game elements were effective in enhancing motivation only when time-based goals rather than attempts-based goals were used. Furthermore, there was no difference in motivational benefits between the use of points and competition in encouraging goal attainment when the difficulty of goals was controlled for. Consequently, time-based goals, alongside both points and competition to encourage players to achieve the goals, were used in this final version of the motorcycle Pelmanism training tool to be evaluated. This meant that the time-based goals were presented competitively, and the number of points awarded for beating the goal was reflective of the level of competitiveness. For example, a goal for completing the round under 1 min that was achieved by 25% of previous players would reward more points than a goal of under 2 min that was achieved by 50% of the players.

The number of Pelmanism rounds that participants would go through during the training session was also shaped by findings from Chapter 6. When presented with time-based goals, the study found that players' performance in terms of time taken to complete each round began to plateau from the third round. This suggests that perceptual learning was greatest during the first three rounds. Furthermore, by the third round, players' motivation to continue to the next round significantly dropped. However, the presence of game elements was able to prevent players in the experimental conditions from losing motivation. Based on these findings, I proposed that three rounds of the motorcycle Pelmanism task is a reasonable number for participants to complete without compromising performance and motivation. During the three rounds, participants would have matched a total of 3x12 motorcycle pairs, with an average of 54 attempts per round (based on Experiment 6 means). Although three rounds are shorter compared to the length of Crundall et al.'s (2017) motorcycle Pelmanism training task, there is a strong rationale for shorter training interventions as they often rely on voluntary engagement and drivers are more likely to take part relative to a longer training intervention.

Measuring the impact of training

As the motorcycle Pelmanism training task aims to improve drivers' perceptual skills for spotting motorcycles, the impact of the training on drivers' motorcycle perceptual skills needs to be assessed. For the motorcycle Pelmanism training tool to be effective, drivers need to demonstrate an improvement in perceptual skills after the training. Therefore, drivers' ability to perceive motorcycles before and after the motorcycle Pelmanism training task needs to be measured and compared against that of a control group who completed a pedestrian Pelmanism task. The ability to perceive motorcycles can be measured using the perceptual tasks in Chapter 3, namely the VR hazard prediction task, visual search task and T-junction task. Although no differences between dual drivers and car drivers were found, these tasks have previously been used to test hazard perception skills (e.g., Crundall, Humphrey, & Clarke, 2008; Sanocki et al., 2015; Ventsislavova & Crundall, 2018). Therefore, the three tasks can still be used to assess whether drivers improved in their ability to perceive motorcycles after completing the motorcycle Pelmanism training task.

The VR hazard prediction task can assess whether the motorcycle Pelmanism training task improves hazard perception ability for motorcycles, by comparing trained and control participants' ability to predict motorcycle hazards. In the motorcycle clips, the conflicting motorcycle would always be visible for a brief duration immediately prior to the occlusion point. This means that drivers need to have looked and processed the conflicting motorcycle in time to predict the motorcycle hazard correctly. Therefore, if the motorcycle Pelmanism task is effective in improving motorcycle perception skills, drivers who completed the motorcycle-specific training should be more accurate at predicting motorcycle hazards than drivers who completed the control training. However, given that accuracy also depends on whether drivers looked at the appropriate direction, the hazard prediction task alone is not sufficient to evaluate the effectiveness of the training because the motorcycle Pelmanism task does not train drivers on scanning behaviours.

In addition to the VR hazard prediction task, the visual search task is also used to evaluate drivers' perceptual skills before and after the motorcycle Pelmanism training task. The visual search task measures how quickly drivers can locate a motorcycle embedded within traffic in an artificial array of traffic images. As the motorcycle can be located in any image within the array, participants need to search the array for features that are representative of motorcycles (Reeder & Peelen, 2013; Yang & Zelinsky, 2009). The motorcycle Pelmanism training task could influence visual search times in two ways. First, it is possible that motorcycle features become more prominent and attract attention more easily after completing the motorcycle Pelmanism training task. However, processing is also required during visual search to verify or identify motorcycles once attention is allocated to a relevant part of the visual field (Reeder et al., 2015). The motorcycle Pelmanism training task could also improve drivers' ability to process motorcycle features quickly and accurately through reducing the threshold or amount of visual information needed to verify the motorcycle. Therefore, if the training is effective, drivers should be quicker to spot the motorcycle within each array after completing the motorcycle Pelmanism training task as compared to pre-training.

As visual search times are inevitably influenced by visual scanning, a third task is used to assess training effects, which minimises the need to scan different locations of the visual field. This third task is the T-junction task and it measures drivers' processing thresholds for detecting and identifying motorcycles when presented in expected locations. The task presents images of the road from the right when a car is emerging from a T-junction and they could be an empty road or could contain a motorcycle or car. This means that a motorcycle may or may not be present in the image, but if it is present, its location is always to the right of the screen. Therefore, participants first have to detect the presence of a vehicle and then process it sufficiently to identify whether it is a car or motorcycle. As the images are presented for a fixed duration of 250ms, detection and identification accuracies can indicate how quickly drivers can process the visual scene to detect and identify a motorcycle respectively. Therefore, if the motorcycle Pelmanism training task is effective in improving perceptual skills for motorcycles, drivers should become faster at processing motorcycles and thus better at detecting and identification motorcycles after completing the training as compared to the control group.

Current study

The aim of this study is to evaluate the effectiveness of the current version of motorcycle Pelmanism task in improving drivers' motorcycle perception. Motorcycle perception is tested using a range of tasks that had been previously reported in the thesis, namely the VR hazard prediction task, visual search task and T-junction task. Participants completed the tasks before training to establish a baseline measure of their motorcycle perception and completed the same tasks again but with different stimuli after training. However, any improvements from the motorcycle Pelmanism training task may only be transitory, particularly if drivers do not take part in regular training sessions. This can be problematic for the real-world implications of the motorcycle Pelmanism training task. Therefore, there is a need to examine the impact of training again after a delay because trained skills may diminish over time. Beanland et al. (2013) raised the concern that the majority of research on driver training interventions lack long-term evaluations of training effects and as a result, the effects of training beyond the immediate post-test is unclear. Consequently, the current study also measured participants' motorcycle perceptual skills four weeks after the training session to determine whether training effects, if any, persisted.

Additionally, a control pedestrian Pelmanism task is used to compare against the motorcycle Pelmanism training task. This is in line with Beanland et al.'s (2013) recommendation to evaluate driver training interventions with randomised controlled designs. Previously, control participants in Crundall et al.'s (2017) study matched fruit pairs but this may not have been a fair comparison to determine the effectiveness of the motorcycle Pelmanism task. This is because matching fruit pairs is not relevant to hazard perception and we cannot be certain whether the observed training effect was simply due to a generic effect of a hazard-related training. Therefore, a pedestrian Pelmanism task is used as the control condition because pedestrians are another category of potential hazards that drivers encounter on the road.

Based on initial evidence that the motorcycle Pelmanism task was effective in improving motorcycle detection rates (Crundall et al., 2017), it is hypothesised that a positive training effect would also be found in this version of the motorcycle Pelmanism task. Although some aspects of hazard prediction such as knowing where to look may not be affected by the motorcycle Pelmanism training task, visual processing is still needed for hazard prediction. Therefore, the motorcycle Pelmanism training task may still have a positive effect on drivers' ability to predict motorcycle hazards on the VR hazard prediction task. With regards to the visual search and T-junction task, the ability to search for, detect and identify motorcycles is dependent on visual processing of motorcycle features. Therefore, participants in the motorcycle Pelmanism condition are hypothesised to be quicker at locating the motorcycle in the visual search task and more accurate at detecting and identifying motorcycles in the T-junction task than participants in the control condition. The improvements in performance on the perceptual tasks are hypothesised to be observed at the two post-training tests – immediately after and four weeks after training.

7.1.2 Methods

Design

A 2x3 mixed design was used. The between-participants variable was exposure to type of Pelmanism training task (motorcycle vs pedestrian Pelmanism) and the within-participants variable was time of test (pre vs post vs delayed). Participants were tested before the training, immediately after and after a period of four weeks. The hazard prediction test was administered in the pre and immediate post-test, whereas the visual search task and T-junction task were administered at all three times of testing. There were four outcome measures in total. For the hazard prediction task, the outcome measure was accuracy in predicting motorcycle hazards. For the visual search task, the outcome measure was time taken to find the image with the motorcycle (for the purpose of this study, participants were not required to search for hackney carriages as in Experiment 2). For the T-junction task, there were two outcome measures, namely detection and identification accuracies. There was also an additional within-participants variable in the T-junction task, which was the type of vehicle shown (car vs motorcycle), resulting in a 2x3x3 design.

Participants

Seventy-one participants took part in this study, of whom 18 were male and 53 were female. Participants were recruited by advertising the study on various platforms such as social media, the university's paid participant panel and the department's credit panel. The only requirement was having at least one year of driving experience. The mean age of the sample was 29.99 years (SD = 12.39) and the mean driving experience was 10.93 years (SD = 11.32). Participants were randomly assigned to the type of Pelmanism training task. There were no significant differences in age or driving experience between training groups ($t(67.6) = -0.89$, $p = .38$; $t(68.3) = -0.99$, $p = .32$ respectively). Participants were not screened for visual skills such as visual acuity or contrast sensitivity.

Materials – Hazard prediction task

The hazard prediction task was used to measure participants' ability to predict motorcycle hazards and consisted of the same set of 10 clips from Experiment 1. The first five clips were used in the pre-training test and the last five clips were used in the immediate post-training test, with one motorcycle hazard clip during each test. The clips were presented in the same order because each clip followed on from the previous to show participants what actually happens. As the clips were developed externally for another project at the university's Transport Research in Psychology group (TRiP), it was not feasible to develop a third set of hazard clips for the delayed post-test due to limited resources of the doctoral project.

Materials – Visual search task

The visual search task was used to measure how quickly participants could detect motorcycles in an artificial array of traffic images. There were three sets of stimuli in total so that a different set of stimuli could be used for each time of testing, randomised across participants. Each set of stimuli consisted of 10 arrays and participants had to search for motorcycles within each array. As there were already 20 arrays from Experiment 2, 10 additional arrays were created in the same way to make the third set of stimuli. In each array, there were 16 images of traffic and only one of the images contained a motorcycle. The location of the target image was assigned randomly across arrays but remained the same across participants. Within each set of stimuli, the order of the arrays was also randomised across participants. The visual search task was coded on Unity as described in Experiment 2.

Materials – T-junction task

The T-junction task was used to measure participants' ability to detect and identify motorcycles accurately when the presence of vehicle is unpredictable, but the location of the vehicles is highly predictable. Similar to the visual search task, there were three sets of stimuli in total, so that one set of stimuli is used at each time of testing in a random order. Each set of stimuli in the T-junction task consisted of 25 empty roads, 20 cars and five motorcycles, created by editing the vehicle on top of the empty roads. Previously in Experiment 3, there was an equal number of motorcycle and car stimuli, and vehicles were presented at three distances (near, mid or far) and at two presentation durations (once at 100ms and once at 250ms). In the current study, the proportion of motorcycle stimuli was reduced to be lower than that of cars so that participants' expectations of encountering a motorcycle stimulus would not be artificially raised. Furthermore, the vehicles were presented only at far distances (approximately 5-7mm in height at full screen) at 250ms because the effects of distance and duration were not variables of interest in the current study. The same empty roads were present across all three sets of stimuli, but the car and motorcycle stimuli, in terms of the vehicle itself and the road it was on, were different. This was to prevent participants not only from recognising a car or motorcycle based on the fact that it was presented previously, but also from associating a particular road with the presence of a vehicle. The T-junction task was programmed in the same way as described in Experiment 3.

Materials – Pelmanism training task

Two versions of the Pelmanism training task were created, one as a control condition where participants matched pedestrian pairs and another as an experimental condition where participants matched motorcycle pairs. The Pelmanism training task consisted of three rounds. Similar to previous designs of the Pelmanism game in this thesis, each round had 24 playing tiles (i.e., 12 matching pairs) and each playing tile had a unique combination of motorcycle or pedestrian and background. Playing tiles for the motorcycle condition were taken from the first three rounds of the 10 rounds of Pelmanism in Chapter 6. Playing tiles for the pedestrian condition were taken from the pedestrian round in Chapter 5, and two additional sets of playing cards were created by re-allocating the same pedestrian pairs to different backgrounds.

The Pelmanism training task was coded and run on Unity. Before each round, participants were presented with difficult, medium and easy time-based goals and received 12, 6 and 2 points respectively (see Figure 7.1 for an example). If participants failed to beat the easy goal, they would receive 0 points. The points system was adjusted slightly as compared to previous experiments (6 and 7), to further incentivise participants to perform well. For instance, the number of points diminished exponentially as the goals got easier and participants would not receive any 'participation' points for failing the easy goal. This is in contrast to previous experiments where participants who failed to beat the easy goal still received 2 points. The difficulty of the goals reflected the percentage of players that had achieved this goal previously and they were calculated from the mean, upper and lower quartiles of participants' time taken for the first three rounds previously in Experiment 7. The final goals used are shown in Table 7.1. A stopwatch and counter were used to record the time taken and number of attempts needed to match all pairs, which were both visible during the game. Similar to the experimental conditions in Experiment 7, the colour of the stopwatch changed according to participants' performance in relation to the targets to provide feedback during the game.

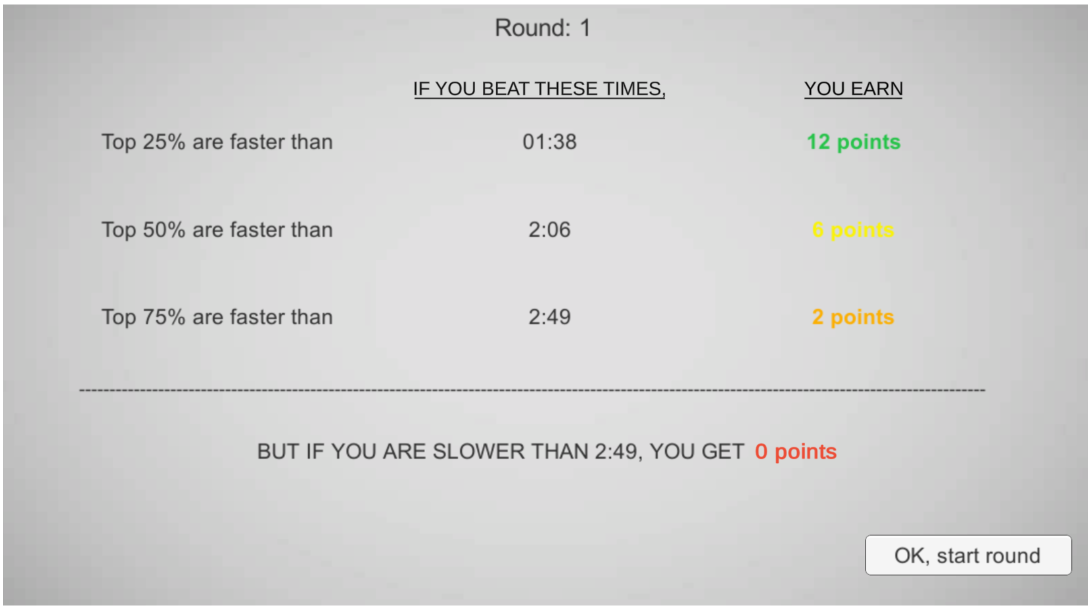

Figure 7.1: Screenshot of goals and rewards shown to participant during the motorcycle and pedestrian Pelmanism tasks

Table 7.1: Goals used in the present study, calculated from time taken by players in previous experiment

Round	Faster than top 25% (12 points)	Faster than top 50% (6 points)	Faster than top 75% (2 points)
1	1:38	2:06	2:49
2	1:19	1:39	2:07
3	1:19	1:35	2:03

Procedure

All participants gave informed consent prior to the start of the study and were informed again that they had to return to the lab for the second session. Participants were randomly allocated to the pedestrian Pelmanism or motorcycle Pelmanism training condition. In the first session, participants completed the three perceptual tasks as pre-test, followed by three rounds of Pelmanism, and then the three perceptual tasks again as a post-training test immediately after. The order of the tasks was counterbalanced across tests. Four weeks later, participants completed a delayed post-training test consisting of the visual search task and T-junction task. The procedure for each task is described in Chapter 3 of the thesis.

Participants were told that they would be playing a game of Pelmanism consisting of three rounds upon completion of the pre-test during the first session. An example of a matching car pair was shown on the screen alongside the instructions to demonstrate that the background and size of motorcycle/pedestrian within each pair was unique. Participants were informed that there would be different goals reflecting different levels of difficulty and they should decide which goal to aim for (Figure 7.1). At the end of each round, participants were asked how motivated they were to continue to the next round and responded on a scale of 1-10 using the keyboard. Participants' actual performance in terms of time taken and number of attempts, as well as performance in relation to the targets and the points scored were shown at the end of each round and game similar to Experiments 6 and 7.

Statistical analysis

All data were analysed using the statistical programming language R (R Core Team, 2020) but separate analyses were conducted for each measure – hazard prediction accuracy, visual search times, detection accuracy and identification accuracy. The analyses are described and reported in the corresponding sections below.

7.1.3 Results

VR Hazard prediction task – Hazard prediction accuracy

Motorcycle clips were the critical trials and only responses on the motorcycle clips were analysed and reported here. This is because there is only one critical trial each in the pre- and immediate post-test and it is not valid to compare the average accuracy of one motorcycle clip to the average accuracy of four non-motorcycle clips. Hazard prediction accuracy for motorcycle trials was thus analysed using a generalised linear model with a binomial link without random effects because there was only one measurement point per participant. As participants were either correct or incorrect (0 or 1), mean proportion of correct responses ranged from 0 to 1. A separate analysis was conducted on the non-motorcycle clips to examine whether the motorcycle Pelmanism task had an effect on hazard prediction accuracy of non-motorcycle hazards. Two participants were excluded from the analysis because they did not complete the VR hazard prediction task due to motion sickness and one participant was excluded due to researcher error.

The mean accuracy for predicting motorcycle hazards before and after each type of Pelmanism training is shown in Figure 7.2. Pelmanism training type was not a significant predictor of accuracy (χ^2 (1) = 1.01, p = .31; *OR* = 1.50 [95% CI: 0.68, 3.37]) but time of test was significant (χ^2 (1) = 6.67, p = .01; *OR* = 2.85 [95% CI: 1.28, 6.67]). Although the means in Figure 7.2 suggest that accuracy in predicting motorcycle hazards was higher after exposure to a motorcycle Pelmanism training task compared to before, this was not significant (χ^2 (1) = 1.43, p = .23; *OR* = 2.73 [95% CI: 0.53, 14.7]).

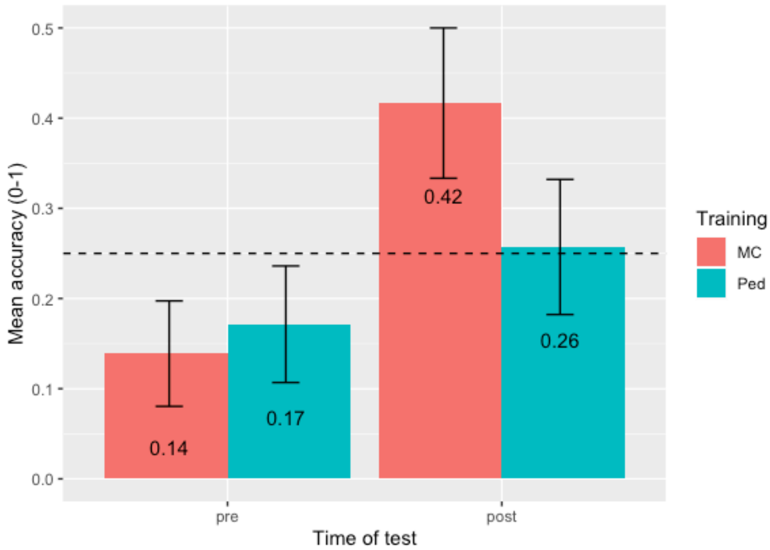

Figure 7.2: Mean accuracy for predicting motorcycle hazards at pre and post-test, according to type of training received, with +/-1 SE and chance performance at 0.25 (dashed line)

Analysing non-motorcycle clips, the time of test was also significant (χ^2 (1) = 16.23, $p < .001$; *OR* = 2.58 [95% CI: 1.62, 4.20]) but there was no main effect nor interaction involving the type of Pelmanism training (χ^2 (1) = 0.33, p = .56; *OR* = 0.88 [95% CI: 0.55, 1.38] and χ^2 (1) = 0.63, p = .43 respectively; *OR* = 0.68 [95% CI: 0.26, 1.76]). Figure 7.3 shows the mean accuracy for predicting non-motorcycle hazards before and after each type of Pelmanism training.

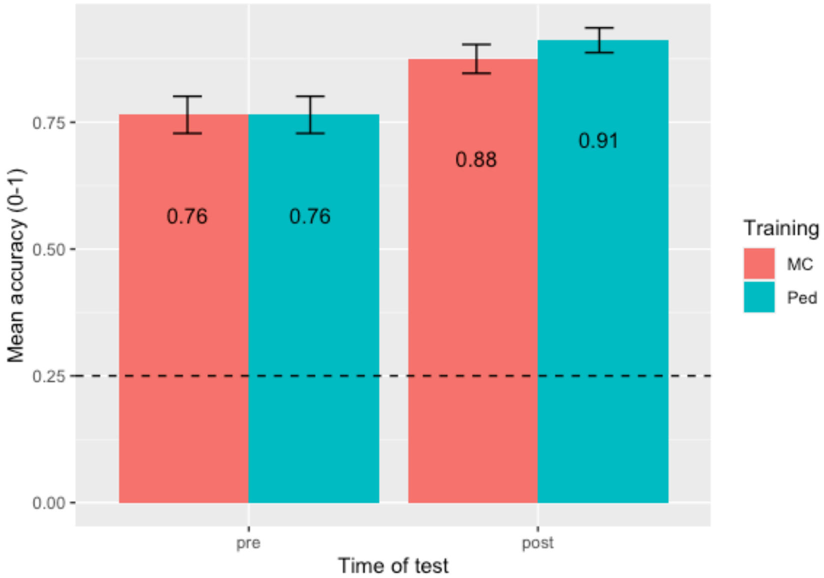

Figure 7.3: Mean accuracy for predicting non-motorcycle hazards at pre and post-test, according to type of training received, with +/-1 SE and chance performance at 0.25 (dashed line)

Visual search task – Search times

Visual search times were analysed using linear mixed effects model and likelihood ratio tests as described in Chapter 2. The full model with the maximal random effects structure is shown below. The inclusion of the random effects improved model fit compared to a general linear model without random effects (refer to Appendix I).

Time taken ~ Pelmanism training type + Time of test + Pelmanism training type:Time of test + (1|Participant) + (1|Array)

Reference/Baseline categories of fixed effects: Training type – pedestrian Pelmanism, time of test – pre-test

Figure 7.4 shows the mean time taken to find the image with the motorcycle before, immediately after and four weeks after completing the motorcycle or pedestrian Pelmanism training task.

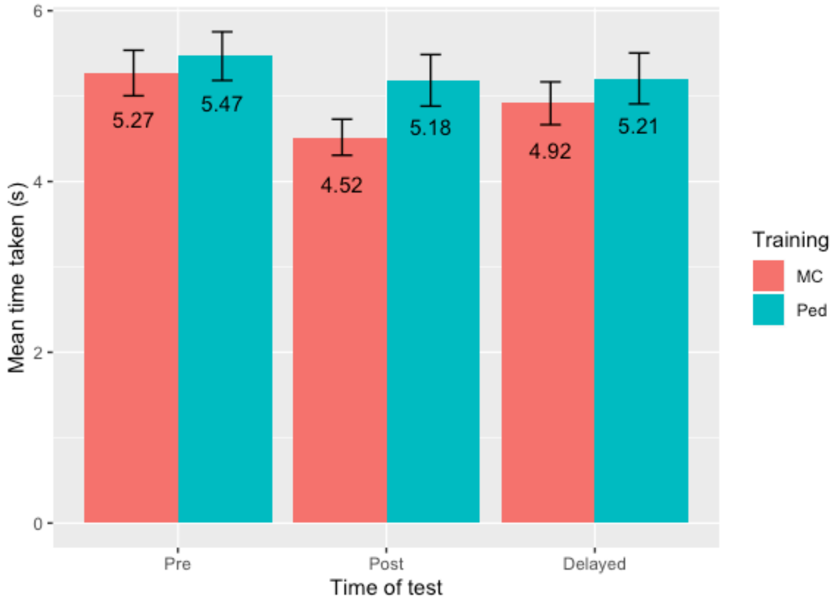

Figure 7.4: Mean time taken to find target image with +/-1 SE at each time of testing according to training type

A logarithm transformation was performed on time taken in the model because the assumption of normality of residuals was violated. Likelihood ratio tests found that the main effect of training type was not significant (χ^2 (1) = 1.79, p = .18; β = -0.063 [95% CI: -0.16, 0.030]). Time of testing was significant (χ^2 (2) = 10.48, p = .0053). Participants took significantly longer to detect motorcycles before training (M = 5.4s, SD = 5.1) than immediately after training (M = 4.8s, SD = 4.8; t (1959.8) = -3.16, p = .0016; β = -0.098 [95% CI: -0.16, -0.037]). The difference in mean time taken before training and after four weeks was not significant after correction for multiple contrasts (M = 5.1s, SD = 5.0; t (1988.0) = -2.19, p = .029, α = .025; β = -0.069 [95% CI: -0.13, -0.0072]). The interaction between type of Pelmanism training and time of testing was not significant (χ^2 (2) = 1.37, p = .50; $\beta_{post\text{-}test}$ = -0.041 [95% CI: -0.16, 0.081] and $\beta_{delayed}$ = 0.032 [95% CI: -0.092, 0.16]).

T-junction task – Detection accuracy

Data from the T-junction task were analysed using linear mixed effects model with a binomial link, similar to Experiment 3, and likelihood ratio tests (refer to Chapter 2). The full model for detection accuracy with the maximal random effects structure is shown below.

The inclusion of random effects improved model fit compared to a general linear model without random effects (refer to Appendix I).

Detection accuracy ~ Pelmanism training type + Time of test + Vehicle shown + Pelmanism training type:Time of test:Vehicle shown + (Time of test|Participant) + (1|Stimulus)

Reference/Baseline categories of fixed effects: Training type – pedestrian Pelmanism, time of test – pre, Vehicle shown – car

The mean detection accuracy for car and motorcycle trials according to type of Pelmanism training at each time of testing is shown in Figure 7.5.

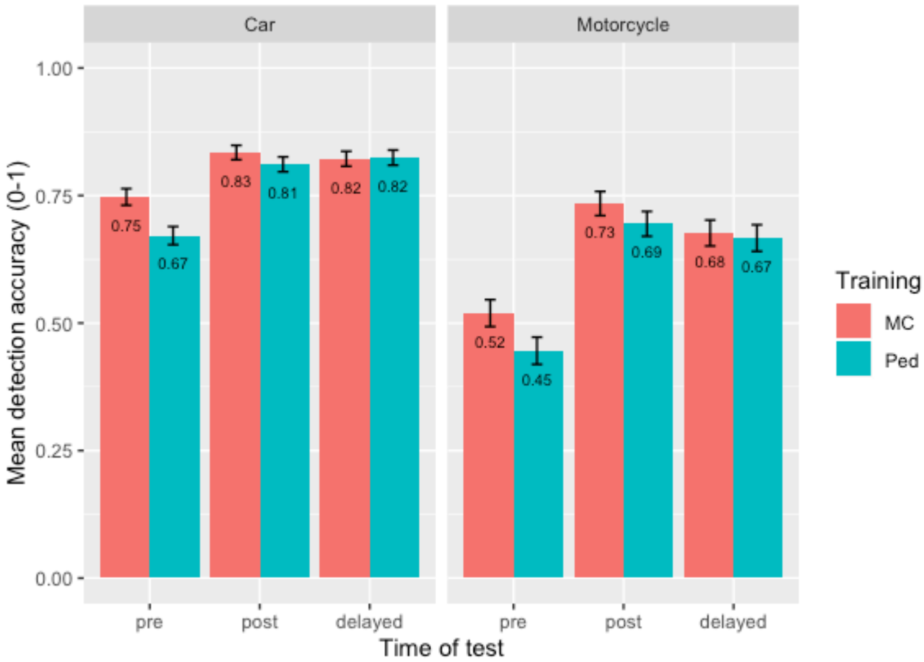

Figure 7.5: Mean detection accuracy with +/-1 SE at each time of testing according to vehicle shown and training type

Likelihood ratio tests found that the type of Pelmanism training was not significant (χ^2 (1) = 0.34, p = .56; *OR* = 1.16 [95% CI: 0.70, 1.92]). Time of testing was significant (χ^2 (2) = 66.16, p < .001). Detection accuracy was significantly lower before the training (M = 0.63, SD = 0.48) as compared to post-test immediately after training (M = 0.79, SD = 0.41; z = 9.89, p < .001; *OR* = 3.20 [95% CI: 2.54, 4.03]) and after four weeks (M = 0.77, SD = 0.42; z

= 8.04, p < .001; *OR* = 2.58 [95% CI: 2.05, 3.25]). The main effect of vehicle shown was also significant (χ^2 (1) = 13.66, p < .001), with cars (M = 0.78, SD = 0.41) detected more accurately than motorcycles (M = 0.62, SD = 0.49; z = -3.84, p < .001; *OR* = 0.34 [95% CI: 0.20, 0.59]). The interaction between time of test and vehicle shown was significant but there were no significant interactions involving type of Pelmanism training (see Table 7.2). Across both types of Pelmanism training, motorcycle detection improved significantly from pre-training (M = 0.48, SD = 0.5) to immediate post-test (M = 0.71, SD = 0.45; z = 3.1, p = .002).

Table 7.2: Summary of interaction analyses for detection accuracy (below significance threshold of .05)*

Interaction	**Df**	χ^2	p	***OR* [95% CI]**
Training type x Vehicle shown	1	0.42	0.52	1.10 [0.83, 1.46]
Training type x Time of test	2	5.59	.061	post-test: 0.71 [0.46, 1.10] delayed: 0.59 [0.38, 0.91]
Time of test x Vehicle shown	2	9.42	.0090*	post-test: 1.74 [1.23, 2.47] delayed: 1.34 [0.95, 1.89]
Training type x Time of test x Vehicle shown	2	0.55	.76	post-test: 1.28 [0.64, 2.56] delayed: 1.21 [0.61, 2.39]

T-junction task – Identification accuracy

Analysis of identification accuracy was conducted on trials where participants correctly detected a vehicle. The full model for identification accuracy with the maximal random effects structure is shown below. The inclusion of random effects improved model fit compared to a general linear model without random effects (refer to Appendix I).

Identification accuracy ~ Pelmanism training type + Time of test + Vehicle shown + Pelmanism training type:Time of test:Vehicle shown + (1|Participant) + (1|Stimulus)

Reference/Baseline categories of fixed effects: Training type – pedestrian Pelmanism, time of test – pre, Vehicle shown – car

The mean identification accuracy of motorcycles and cars by type of Pelmanism training at each time of testing is shown in Figure 7.6.

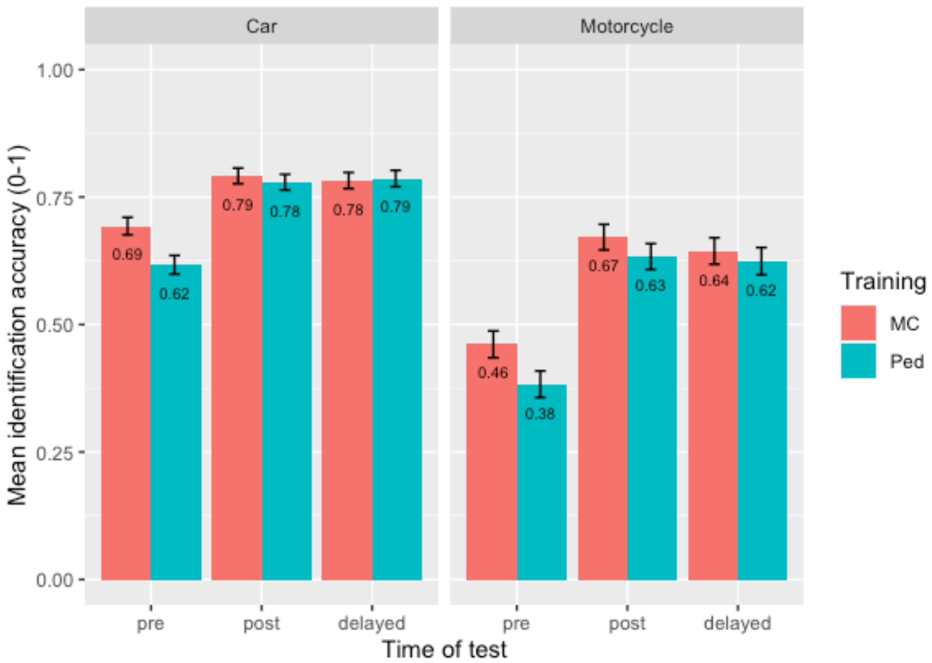

Figure 7.6: Mean identification accuracy with +/-1 SE at each time of testing according to vehicle shown and training type

Likelihood ratio tests found that the type of Pelmanism training was not significant (χ^2 (1) = 0.50, p = .48; *OR* = 1.20 [95% CI: 0.73, 1.97]). Time of testing was significant (χ^2 (2) = 24.77, p < .001). Identification accuracy was significantly lower before training (M = 0.58, SD = 0.49) as compared to post-test immediately after training (M = 0.74, SD = 0.44; z = 4.17, p < .001; *OR* = 1.84 [95% CI: 1.38, 2.44]) and after four weeks (M = 0.73, SD = 0.44; z

= 4.76, $p < .001$; *OR* = 2.09 [95% CI: 1.54, 2.84]). The main effect of vehicle shown was also significant (χ^2 (1) = 3.95, $p = .047$), with cars (M = 0.78, SD = 0.41) being identified more accurately than motorcycles (M = 0.62, SD = 0.49; z = -2.06, $p = .039$; *OR* = 0.59 [95% CI: 0.36, 0.98]). None of the interactions were significant (see Table 7.3).

Table 7.3: Summary of interaction analyses for recognition accuracy

Interaction	Df	χ^2	*p*	*OR* [95% CI]
Training type x Vehicle shown	1	0.94	.33	1.30 [0.78, 2.18]
Training type x Time of test	2	1.12	.57	post-test: 0.75 [0.42, 1.32] delayed: 1.00 [0.55, 1.83]
Time of test x Vehicle shown	2	2.77	.25	post-test: 1.02 [0.56, 1.85] delayed: 1.69 [0.88, 3.25]
Training type x Time of test x Vehicle shown	2	0.17	.92	post-test: 1.01 [0.31, 3.34] delayed: 1.30 [0.35, 4.80]

7.1.4 Discussion

The present study examined the effectiveness of the motorcycle Pelmanism task in improving motorcycle perception, measured using the hazard prediction task, visual search task and T-junction task. The results indicate that the motorcycle Pelmanism training task used in the current study did not improve drivers' motorcycle perceptual skills on the three tasks. The lack of training effect on participants' motorcycle perceptual skills was unexpected because there is previous evidence that subordinate level training improves perceptual skills and increases sensitivity towards diagnostic features of trained objects (Crundall et al., 2017; van der Linden et al., 2014). A brief inspection of participants' motivation ratings indicates that participants were similarly motivated in each Pelmanism training condition, therefore differences in motivation and engagement levels between the two Pelmanism conditions could not account for the lack of motorcycle training effect[2].

Participants who completed a motorcycle Pelmanism training task were not better able to predict motorcycle hazards than participants who completed a pedestrian Pelmanism task. This suggests that the motorcycle-specific training developed in this thesis had little effect on participants' motorcycle hazard perception ability, measured by prediction accuracy rather than response times. One account for the lack of training effect on hazard prediction ability could be because the VR hazard prediction task is dependent on both the looking and processing stage of hazard perception. Drivers need to have looked at the right locations at the right time and then to identify the hazard in time, but the focus of the motorcycle Pelmanism training task is not on training participants where to look. The scarcity of VR motorcycle clips could also have contributed to the lack of training effect as there may have been insufficient trials to accurately measure participants' hazard prediction ability for motorcycles, resulting in high levels of variability in the means. This could explain why the interaction between type of training and time of test did not reach statistical significance even though the means suggest that accuracy in predicting motorcycle hazards was higher after completing the motorcycle Pelmanism task compared to the pedestrian Pelmanism task. There were also no motorcycle-specific improvements in the visual search nor T-junction task for trained participants. Participants who completed the motorcycle Pelmanism training task were not quicker to search for motorcycles, detect nor identify motorcycles more accurately in the T-junction task than participants who completed the control Pelmanism task. This suggests that the motorcycle Pelmanism training task used in this study was not successful in increasing participants' sensitivity to motorcycle features, such that motorcycles

[2] The mean motivation rating of participants in the motorcycle Pelmanism condition and pedestrian Pelmanism condition was 7 (S.D. = 2.4) and 8 (S.D. = 1.7) respectively, which did not differ significantly (χ^2 (1) = 3.5, p = .06).

are more likely to capture attention during visual search, nor lower their processing thresholds for detecting and identifying motorcycles.

Although there were no training effects arising from the motorcycle Pelmanism task, participants from both training conditions improved from pre to post-training across all motorcycle perception measures. The most immediate explanation is that practice effects were stronger than any training effect by the motorcycle Pelmanism task. As a different set of stimuli was used at each point of testing and counterbalanced across participants, repeated exposure to test stimuli could not have led to practice effects. Instead, practice effects in terms of task learning may have occurred, whereby participants became familiarised with the task demands. While this may be a contributing factor, there could be other reasons for the general improvement observed in both conditions. As there were proportionately more car stimuli than motorcycle stimuli in the T-junction task, we would expect that car detection would improve at a quicker rate because participants had more practice with car detection as compared to motorcycle detection. However, results from the T-junction measure revealed that both Pelmanism training conditions led to a greater immediate improvement in motorcycle detection than car detection. This suggests that the general improvement cannot be explained solely by practice effects. In contrast, there may be a general priming effect of subordinate level matching task in both Pelmanism training tasks, which will be further elaborated on in the following paragraph.

Participants in both Pelmanism conditions may have developed the same underlying skill that benefitted performance on the three tasks. The pedestrian and motorcycle Pelmanism tasks required participants to discriminate between features of each stimulus to make a match. For example, participants may remember 'man in blue shirt with stroller' or 'black motorcycle with two headlights'. By encouraging participants to pay attention to high level features or details in the stimuli, the global precedence effect (Navon, 1997) may have been attenuated as participants were primed to attend to local features alongside global information. Normally, the tendency is to process low spatial frequency targets such as cars before high spatial frequency targets, which is one of the reasons why drivers tend to overlook motorcycles but not cars (e.g., Pammer et al., 2018). This tendency is especially evident when motorcycles are surrounded by wide objects of low spatial frequency (Rogé et al., 2010). It is possible that the subordinate level matching in both Pelmanism conditions may have primed participants to direct relatively equal amounts of attention to both global and local features in order to process the high level of details during matching. As a result, even though participants improved in both car and motorcycle detection, they may have improved significantly more in motorcycle detection than car detection because the tendency to process low over high spatial frequency objects was inhibited after subordinate-level Pelmanism training. Therefore, subordinate-level matching could have contributed to the improvements in motorcycle detection in both training conditions. Following this reasoning, the primed effect resulting from the subordinate-level matching did not persist when participants were tested again after four weeks, likely because participants did not complete another Pelmanism game before the delayed post-test.

Overall, the present findings do not support Crundall et al.'s (2017) findings, who reported a positive training effect for participants who completed the motorcycle Pelmanism task. One potential reason may be due to the difference in number of rounds included in the motorcycle Pelmanism training task. Participants in the present study only completed three rounds of the motorcycle Pelmanism task whereas participants in Crundall et al.'s (2017) study completed five rounds. Three rounds were decided based on the findings in Chapter 6 that the game elements were effective from the third round and participants' performance also began to plateau from approximately the third round. It is possible that three rounds of the motorcycle Pelmanism task, while beneficial for appealing to drivers to take part, was insufficient to produce a change in motorcycle perception. As the use of a Pelmanism-based training task targeted at improving motorcycle perception is still a relatively recent concept, its effectiveness should not be ruled out until further research is conducted. More research is needed to understand the impact of varying different features in the motorcycle Pelmanism training task such as the number of rounds. Future studies could compare the effectiveness of a motorcycle Pelmanism training task with three or five rounds to determine whether the absence of a motorcycle-specific training effect in the current study is due to a difference in the number of rounds.

In conclusion, the present motorcycle Pelmanism training tool shaped by the findings of this thesis was not effective in improving drivers' motorcycle perception after a single session of training. General improvements observed in both the motorcycle and pedestrian Pelmanism tasks were attributed to factors including practice effects and the influence of subordinate-level matching on visual processing hierarchy and the global precedence effect. However, there were no motorcycle-specific training effects resulting from the motorcycle Pelmanism task, which could be due to the use of fewer rounds compared to Crundall et al.'s (2017) version of the motorcycle Pelmanism training task. Despite this, the motorcycle Pelmanism training task remains a promising training approach for improving motorcycle perception while addressing driver engagement based on the literature. Continued research is still needed to advance our understanding of such as training approach and how different elements of the motorcycle Pelmanism training task could affect its impact. As mentioned in the discussion, a logical next step for research would to be investigate the possibility that the number of rounds influences the effectiveness of the motorcycle Pelmanism training task.

Chapter 8 – General Discussion

8.1 Overview of thesis

In the first chapter, a review of the literature and collision statistics found that motorcyclists are worryingly likely to be involved in collisions where another driver looked but failed to see the motorcycle. However, there is some evidence that drivers who also ride motorcycles (i.e., dual drivers) are less susceptible to committing these perceptual errors around other motorcyclists. Therefore, the experiments reported in this thesis were designed to investigate how perceptual failures leading to LBFTS collisions with motorcycles could be prevented through two main research questions – 1) Do dual drivers have better motorcycle perceptual skills compared to car drivers and 2) Can car drivers' perceptual skills for motorcycles be improved using a game-based training tool known as the motorcycle Pelmanism task. Experiments 1, 2 and 3 compared the motorcycle perception abilities of dual drivers and car drivers across a range of perceptual tasks. The aim was to examine whether familiarity developed from real-world experience with motorcycles enhances dual drivers' motorcycle perceptual skills in terms of hazard prediction ability, visual search times, detection and identification abilities. However, no motorcycle-specific perceptual advantage for dual drivers were found on the three tasks used.

The fourth chapter highlighted the limitations of existing hazard perception training interventions, specifically the lack of emphasis on hazards involving motorcycles and engaging elements to encourage participation. In contrast, serious games have the potential to achieve learning outcomes while engaging learners. A promising motorcycle perceptual training task based on a game of Pelmanism was identified and the remaining experiments were designed to validate, further develop and evaluate the motorcycle Pelmanism task as a training tool. Experiments 4 and 5 examined whether the motorcycle Pelmanism matching task was sensitive to real world experience and familiarity with motorcycles. Comparing dual drivers' and car drivers' performance in terms of time taken and number of attempts, dual drivers consistently required fewer attempts to complete the motorcycle Pelmanism task. This suggests that the motorcycle Pelmanism task does tap into relevant skills that are associated with motorcycle familiarity. Having examined the sensitivity of the motorcycle Pelmanism task, the next two experiments were conducted to determine the optimal number of rounds to include in the training and the effectiveness of specific goals, points and competition as game elements. The final experiment then evaluated the effectiveness of the motorcycle Pelmanism training task developed throughout Experiments 4-7, using perceptual measures reported in the initial experiments.

8.2 Is there a dual driver advantage in motorcycle perception?

8.2.1 Summary of findings

The first aim of the thesis was to examine whether dual drivers have a perceptual advantage for motorcycles compared to car drivers and in doing so, to identify tasks that could be used to evaluate the motorcycle Pelmanism training tool in the final study. Three experiments were conducted in Chapter 3 to investigate the effect of motorcycle experience on drivers' ability to predict hazards involving motorcycles, to spot motorcycles in a visual search array, to detect and identify a motorcycle at a T-junction. These experiments examined drivers' motorcycle perception in terms of whether they looked at and then processed the motorcycle because failures at either stage would result in failure to perceive the motorcycle (Crundall, Clarke et al., 2008).

The first experiment examined drivers' ability to predict motorcycle hazards in a 360° virtual reality environment. Previous research suggests that dual drivers have better hazard perception ability than car drivers, but previous studies have only tested general hazard perception ability for a range of common hazards (Horswill & Helman, 2003). The rationale for this experiment was to determine whether dual drivers have a motorcycle-specific advantage in hazard perception. According to the literature reviewed, response times measure used in traditional hazard perception tests is often confounded by drivers' perceived level of risk in each situation (Borowsky & Oron-Gilad, 2013; Crundall, 2016). Measuring accuracy in predicting hazards is advantageous over response times because the hazard prediction test directly measures whether drivers noticed the hazard. However, the results of the study showed that dual drivers were not more accurate at predicting motorcycle hazards than car drivers. As the clips were cut such that the conflicting motorcycle was visible briefly immediately prior to the occlusion, this likely meant that dual drivers did not have an advantage in terms of looking in the appropriate location and processing it quickly enough to have notice the conflicting motorcycle. This was inferred from drivers' accuracy in predicting the motorcycle hazard and the hazard prediction test could certainly be improved in future studies by including an eye-tracking measure to determine whether drivers fixated on the motorcycle.

The second experiment compared how quickly dual drivers and car drivers could spot motorcycles in a visual search task. Categorical search requires top-down knowledge of features that are representative of the target category to help direct attention during search (Hollingworth, 2012). An attentional set is activated based on knowledge of these features and is used as a search template (Reeder & Peelen, 2013). As dual drivers have greater exposure to motorcycles, dual drivers are likely to be more familiar with motorcycle features than car drivers, which could aid categorical search. In contrast to my hypothesis, the results found that dual drivers were not quicker at spotting motorcycles in an array of traffic images than car drivers. This suggests that motorcycles were not more likely to attract dual drivers' attention than car drivers' attention, despite dual drivers being more familiar with motorcycles. As the motorcycle target could be located anywhere in the search array, drivers could not rely on their schemas to guide their attention during visual search but had to scan the entire array. The time taken to detect a motorcycle may therefore largely consist of time taken to scan the array. Consequently, the visual processing abilities of drivers may not have been adequately assessed with the visual search task. The third experiment thus examined the processing thresholds for detecting and identifying a motorcycle when it appears at a T-junction, which requires less scanning than an array of traffic images.

The T-junction task in the third experiment presented images of empty roads, cars or motorcycles approaching from the road on the right when taken from the perspective of a car driver emerging from a T-junction. The thresholds for detecting the presence of a motorcycle and identifying it as a motorcycle were examined. By presenting stimuli at a long and short duration, the third experiment aimed to determine whether dual drivers have a lower processing threshold for detecting and identifying motorcycles than car drivers. There is some evidence that dual drivers are more accurate at detecting vehicles in the T-junction task than car drivers (Crundall et al., 2017; Lee et al., 2015), but there is yet to be evidence of a motorcycle-specific advantage at detection or identification for dual drivers. Compared to car drivers, dual drivers have more exposure to motorcycles and thus should be able to detect and identify motorcycles more quickly and accurately. However, the results showed no difference between dual drivers and car drivers in detection and identification accuracy, even when motorcycles were presented at far distances and for short durations.

Interestingly, unlike the three perceptual tasks, the Pelmanism task did show a motorcycle-specific advantage for dual drivers. In Chapter 5, Experiments 4 and 5 compared the performance of dual drivers and car drivers on a motorcycle and pedestrian Pelmanism task. Across both experiments, dual drivers consistently required fewer attempts when matching motorcycle pairs than car drivers, but this advantage was not present when matching pedestrian pairs. The findings indicate that dual drivers were more accurate than car drivers at identifying motorcycle stimuli at a subordinate level, which is in line with previous research on the influence of familiarity on memory accuracy (e.g., Herzmann & Curran, 2011; Wilson et al., 2011). As the Pelmanism task requires drivers to process and encode motorcycles to a sufficient level in order to produce a memory trace, the findings suggest that dual drivers were better than car drivers at perceptually encoding the motorcycle stimuli at a featural level to allow subordinate-level discrimination. However, there is inconclusive evidence on whether dual drivers were quicker at matching motorcycle pairs than car drivers. Therefore, based on the current findings, the familiarity advantage may be demonstrated more robustly through a game of Pelmanism in terms of matching accuracy.

8.2.2 Discussion of findings

Altogether, my findings indicate that dual drivers do not have a motorcycle-specific advantage on the three out of the four motorcycle-related tasks that compared dual drivers against car drivers. The three perceptual tasks from Experiments 1-3 found no evidence that dual drivers have superior visual processing abilities in terms of looking and processing for motorcycles. This challenges the notion that experience in motorcycle riding makes drivers better at spotting other motorcyclists on the road as compared to drivers without motorcycle riding experience. Previously, researchers have suggested and hypothesised that dual drivers have better perceptual skills for motorcycles than car drivers, based on the premise that knowledge of and familiarity with motorcycles can influence eye movements, attention and processing thresholds (Crundall, Humphrey, & Clarke, 2008; Lee et al., 2015). This has been a long-standing assumption and research examining dual drivers' collision rates with motorcycles may have also fuelled the idea that dual drivers are better at spotting other motorcyclists. For instance, Magazzù et al. (2006) reported that dual drivers were less likely than car drivers to collide with motorcycles and a motorcycle-specific benefit was also reported in close friends and relatives of motorcycle riders who have ridden pillion (Brooks & Guppy, 1990, as cited in Crundall et al., 2012). Researchers may have subsequently overestimated the benefit of familiarity on motorcycle perceptual skills, despite little evidence suggesting that the lower collision rates are due to enhanced perceptual skills for motorcycles. As mentioned in the introduction, there are other top-down and bottom-up factors that contribute to LBFTS collisions with motorcycles, for example expectations and inattentional blindness. Therefore, it is possible that dual drivers' lower collision rates with motorcycles compared to car drivers could be due to other non-perceptual advantages, such as increased sustained attention to the roadway. Future studies could investigate whether there are non-perceptual differences between dual drivers and car drivers that could account for dual drivers' lower collision rates with motorcycles.

Considering that dual drivers outperformed car drivers in the motorcycle Pelmanism task but not in the first three perceptual tasks, it is possible that dual drivers have an advantage only in instances where they have control over the amount of visual processing or effort needed for the motorcycle-related task. For instance, it is unlikely that dual drivers can strategically increase efforts to become better at predicting motorcycle hazards if they had not looked at the right locations at the right time, nor can they dictate how readily motorcycle features stand out when searching for motorcycles. Similarly, dual drivers were unable to control how much visual information is used to detect or identify motorcycles because stimuli were presented at fixed durations. In contrast, dual drivers could self-determine the amount of effort or level of processing during the Pelmanism game, for instance by paying more attention to certain features or details when matching the motorcycle pairs. This could explain why a motorcycle-specific advantage for dual drivers was found in Experiments 4-5 but not in Experiments 1-3. Another example comes from Crundall et al.'s (2012) study in which there were no restrictions on presentation durations. They found that dual drivers were safer and fixated longer on conflicting motorcycles than car drivers. This suggests that dual drivers were sensitive to the level of processing required for motorcycles and consequently put more effort into processing motorcycles by fixating for longer. Yet, dual drivers were not faster than car drivers to first fixate on the motorcycle, which indicates that motorcycles did not capture dual drivers' attention any more than car drivers. Therefore, it is possible that dual drivers are no better at spotting the motorcycle than car drivers when the amount of visual information available is restricted and drivers are unable to regulate the amount of visual processing.

Rather than a motorcycle-specific perceptual advantage, my findings suggest that dual drivers have a general advantage in spotting both motorcycles and cars. Findings from the T-junction task in Experiment 3 showed that even at short presentation durations, dual drivers were better at detecting both vehicles but did not exhibit a motorcycle-specific perceptual advantage. A similar observation was reported by Crundall et al. (2017), who found that dual drivers were more accurate at detecting both cars and motorcycles than car drivers. A close examination of Lee et al.'s (2015) findings also revealed that compared to British drivers, the perceptual advantage exhibited by Malaysian drivers, who have a high level of exposure to motorcycles, emerged in terms of a smaller discrepancy between car and motorcycle detection rather than an enhanced advantage for motorcycles specifically. Overall, it appears that the perceptual advantage for drivers who are familiar with motorcycles arises in the form of enhanced perceptual skills for all vehicles rather than for motorcycles specifically. A likely reason for the general advantage could be that the skills for detecting motorcycles can be transferred easily to detecting other vehicles. Motorcycles are easily overlooked due to their narrow and small outline whereas the other vehicles on the road are almost always larger than motorcycles. Therefore, as drivers' ability to detect motorcycles develops with repeated exposure, by extension their ability to detect any vehicle larger than a motorcycle likely also improves, thus producing the general advantage whereby drivers are better at spotting motorcycles and cars.

8.2.3 Implications and future directions for research

The implication of my findings is that researchers need to re-evaluate the concept that dual drivers have a motorcycle-specific perceptual advantage. Until there is consistent evidence demonstrating this, researchers should refrain from perpetuating the assumption that drivers with motorcycle riding experience have a perceptual advantage for spotting motorcycles compared to drivers who only drive cars. Magazzù et al.'s (2006) study has been a popular reference for justifying a dual driver advantage in motorcycle perception and to my knowledge, has not been replicated. This is alarming because the direction of a research field is easily misguided when it hinges upon on the same few references. Magazzù et al. (2006) analysed car-motorcycle collisions taken from the MAIDS dataset and examined the contributory variables of each collision. They reported that having a motorcycle riding license reduces the risk of causing collisions with motorcycles. To determine if this pattern of statistics can be replicated, an up-to-date analysis of a different sample of car-motorcycle collisions is warranted. Considering the emerging evidence that dual drivers exhibit a general rather than motorcycle-specific perceptual advantage, the updated analysis should also examine dual drivers' likelihood of colliding with a car and colliding with a motorcycle. This could help determine whether a general advantage manifests in collision risks involving both vehicle types. Most crucially, research should aim to establish whether there is a causal link between dual drivers' perceptual advantage and their low collision rates. This is because there may be other factors that reduce dual drivers' likelihood of collision with motorcycles, such as an increased alertness to the presence of motorcycle (see Gershon et al., 2012).

There is also a need to continue testing whether a dual driver perceptual advantage for motorcycles exists and future studies should continue current efforts to examine this proposition. So far, my findings have provided strong evidence against a dual driver perceptual advantage, however the majority of my studies used static images of motorcycles. A visual processing advantage for dual drivers could emerge in studies that involve dynamic clips of motorcycles because of a confluence of other cues, such as motion, which dual drivers may rely on to detect motorcycles. For example, a motorcycle appearing between cars or at lane positions where cars do not conventionally take up might capture drivers' attention exogenously, reducing detection times and increasing detection accuracies. Although one of my studies – the VR hazard prediction task used dynamic clips, more evidence is needed to gain a comprehensive understanding of whether a dual driver perceptual advantage arises in dynamic motorcycle clips. As mentioned previously, the VR hazard prediction task may also benefit from a direct eye-tracking measure to determine if drivers actually fixated on the developing hazard. Future studies could also replace static images with dynamic clips in the T-junction task, similar to that reported in Crundall et al.'s (2012) study. However, instead of recording response times reflecting when drivers thought it was safe to pull out as the authors did, future studies could measure response times for detecting motorcycles and cars. This would inform us whether dual drivers are quicker than car drivers to detect and identify motorcycles in the presence of dynamic cues. Future studies should also measure and control for any driver group differences in visual skills such as visual acuity or contrast sensitivity. This is because impairments in contrast sensitivity and visual acuity may reduce reaction time to hazards and visual acuity may be correlated with collision involvement (see Swan et al., 2019; Treat et al., 1979).

8.3 Development of the motorcycle Pelmanism training task

The second aim of the thesis was to further develop and evaluate a motorcycle-specific training based on the Pelmanism game. Despite the overrepresentation of motorcyclists involved in collisions caused by other drivers, there is a lack of training interventions that focus on improving drivers' hazard perception for motorcycles specifically. In addition, existing training interventions are often viewed as formal interventions for not-so-competent drivers that lack in hazard perception skills. Consequently, drivers are unlikely to engage with hazard perception training interventions voluntarily, because they believe that they are adequately skilled at hazard perception. The motorcycle Pelmanism task is a potential solution that addresses the need for a motorcycle-specific training as well as the lack of appeal for drivers to take part in the training. Crundall et al. (2017) reported that drivers who completed a motorcycle Pelmanism training task were subsequently better at detecting motorcycles in a T-junction task as compared to drivers who completed a control Pelmanism training task. The overall positive finding provides preliminary evidence for the use of a motorcycle-specific and Pelmanism-based training task. Crundall and colleagues (2017) were the first to examine the use of Pelmanism in the context of motorcycle training and this thesis makes an original contribution to the literature by further developing and assessing the concept of a Pelmanism-based motorcycle training task. A series of experiments were run to validate and further develop the motorcycle Pelmanism training task, before finally evaluating its effectiveness.

8.3.1 Validity of a motorcycle Pelmanism-based training task

Summary and discussion of findings

In Chapter 5, Experiments 4 and 5 were conducted to compare the performance of dual drivers and car drivers on a motorcycle Pelmanism task. These experiments provided the first validation of the motorcycle Pelmanism task by assessing its sensitivity to the performance of drivers who are familiar with motorcycles. A motorcycle-specific advantage for dual drivers was found in matching motorcycle pairs. It is important for a training task to demonstrate sensitivity to players' level of familiarity because it ensures that the training task taps into skills that are reflective of real-world experience with and knowledge of motorcycles (see Graafland et al., 2012). The presence of a dual driver advantage therefore indicates that the motorcycle Pelmanism-based training task is tapping into relevant skills that drivers with real world experience with motorcycles possess, and further strengthens Crundall et al.'s (2017) reported training effect.

Implications and future directions for research

The findings indicate that performance on a game of Pelmanism, particularly in terms of matching accuracy, is influenced by familiarity with the stimuli used. An implication of this finding is that a Pelmanism game involving same-category stimuli may be used to assess familiarity levels, which is distinct from perceptual expertise. Perceptual experts are usually characterised by their superior abilities in detecting, categorising and identifying objects of their expertise, but someone who is familiar with a certain category of objects may not necessarily be a perceptual expert. For instance, even though dual drivers have greater exposure to and more experience with motorcycles than the average car driver, they still did not exhibit perceptual expertise for motorcycles in the three perceptual tasks reported in Experiments 1-3. Therefore, while dual drivers are more familiar with motorcycles than the average car driver, they may not yet be considered perceptual experts. Real-world perceptual experts may only refer to a small group of individuals who have had extensive experience in their domain of expertise whereas familiarity is likely to be more reflective of the general population than perceptual experts. A Pelmanism-based task may therefore be useful for assessing familiarity in individuals who are not yet perceptual experts.

One direction for future research is to continue investigating how performance on the Pelmanism task is facilitated by familiarity or knowledge of the stimuli category used in the game. In Experiment 5, a novel approach was taken in an attempt to examine how familiarity and real-world experience impacts performance on the Pelmanism task, by defining three distinct stages based on players' actions during a game of Pelmanism. Future studies could capitalise on my study and develop a proper psychological model of how players engage with the Pelmanism game, by taking into account how players' strategies and moves change throughout the game. For instance, a game-theory approach could be adopted to make inferences about players' strategies by examining each click in the context of previous clicks. As my data were collected online, factors such as the need for servers and cloud storage limited the ability to record each click throughout the game in my experiments. In addition, the size of the screen and conditions in which participants completed the task could not be accounted for in online testing. Future studies could use face-to-face testing and improve on the method of data collection to record the location and timestamp of each click. This would then allow researchers to analyse each click and the number of revisits for each stimulus or playing tile.

8.3.2 Importance of game elements

Summary and discussion of findings

To address the need to appeal to and encourage drivers to take part in a driver training intervention, I aimed to improve upon the motivational features of the motorcycle Pelmanism training task in Chapter 6. The effects of goals, difficulty-dependent points and competition as game elements were examined in two experiments. In one experiment, the goals represented number of attempts whereas in the other experiment, the goals represented time taken to complete each round. According to the results, time-based goals together with competition or points that reward players according to the difficulty level of the goals were both effective in motivating players. Players remained more motivated across more rounds in the presence of time-based goals, regardless of whether points or competition were used to encourage goal attainment, than when asked to do their best in the control condition. However, this was not observed when goals representing number of attempts were used. Accuracy-based goals may not have been as effective at motivating players as time-based goals because of the discrete nature of attempt counts in comparison to the continuous nature of time taken. In other words, accuracy-based goals may have resulted in intermittent bursts of motivation whenever participants made an attempt. In contrast, the sense of urgency resulting from time-based goals accumulates throughout the entire game, which would have a continual effect on players' motivation. Although the Pelmanism can be used as it is (e.g., Crundall et al., 2017), my findings provide evidence that game elements can be used as motivating features to enhance drivers' motivation to take part in the Pelmanism-based training task. Therefore, there is potential for incorporating additional game elements into serious games and game-based training.

Implications and future directions for research

There is a need for effective and appropriate motivational features to be incorporated into driver training interventions, in order to engage and appeal to drivers. As driver training interventions rely on drivers' voluntary participation, it is important for the training intervention to attract their interest and increase their willingness to participate. The potential for incorporating additional game elements is also applicable to serious games or game-based training. For instance, although the Pelmanism task on its own is already a game, the matching task can become too repetitive and drivers may lose interest after a few rounds, as demonstrated in the control condition without game elements in Experiments 6 and 7. Therefore, by incorporating game elements into the Pelmanism-based training task, drivers' interest in the training can be sustained for longer. As noted in a report by Pressley et al. (2017), the use of incentives should encourage participation and overcome driver apathy towards the training intervention. Although game elements are promising motivational features for training interventions, it is important to highlight that they should not be implemented in a one-size-fits-all manner. Researchers or developers of driver training interventions should test the effects of any game elements in the context they are to be used in before implementing them. For instance, Experiments 6 and 7 showed that time-based goals but not accuracy-based goals were suitable for the motorcycle Pelmanism task. Therefore, to develop a training intervention with participation incentives, researchers should first identify potential game elements from the literature, before empirically examining how each game element affects participants' motivation and performance on the training.

8.4 Was the motorcycle Pelmanism training task effective?

8.4.1 Summary and discussion of findings

Finally, Chapter 7 evaluated the effectiveness of a version of the motorcycle Pelmanism training task developed in this thesis in improving drivers' motorcycle perception. To evaluate the motorcycle Pelmanism training task, the pre- and post-training measures need to be capable of measuring drivers' perceptual skills for motorcycles. Several perceptual tasks had already been developed to examine drivers' motorcycle perception in Experiments 1-3, namely the hazard prediction task, visual search task and T-junction task. The rationale was to first assess dual drivers' performance of these perceptual tasks before using them to measure the impact of the motorcycle Pelmanism training task. If trained participants performed better on the same perceptual tasks that had demonstrated sensitivity to drivers who are familiar with motorcycles, this would enhance the credibility of the motorcycle Pelmanism training task. However, according to Experiments 1-3, dual drivers did not perform better on the three perceptual tasks than car drivers. Although the tasks did not find a dual driver advantage for motorcycles, they are still useful for assessing motorcycle perception and can be used to measure improvements before and after training for several reasons. First, there is little evidence that a dual driver perceptual advantage actually exists. As a result, it would be illogical to invalidate the tasks' abilities to assess motorcycle perception simply because they did not demonstrate an effect that may not even exist. Evidence of a dual driver advantage can be sufficient but is not necessary to demonstrate that these tasks are capable of measuring motorcycle perception. Secondly, hazard prediction ability, visual search skill, detection and identification abilities all contribute to understanding how drivers perceive hazards. Therefore, it is still relevant to examine how drivers perform on these tasks to measure the impact of the motorcycle Pelmanism training task. Lastly, even if other perceptual tasks were to be used to measure the impact of training, there is still no guarantee that a motorcycle-specific advantage for dual drivers would be found. In light of these reasons, I chose to use the same three perceptual tasks reported in Chapter 3 to evaluate the effectiveness of the version of motorcycle Pelmanism training task presented in this thesis. Drivers completed either three rounds of the motorcycle Pelmanism task as training or three rounds of a control pedestrian Pelmanism task, and their motorcycle perceptual skills on the three perceptual tasks before and after training were compared. Overall, the results showed that three rounds of the current motorcycle Pelmanism task did not improve drivers' ability to predict motorcycle hazards, search for, detect or identify motorcycles as hypothesised.

The lack of training effects in my study contradicts previous findings from Crundall et al. (2017), who reported that drivers who completed the motorcycle Pelmanism training task subsequently improved in detecting motorcycles. One possible reason is that the training effect is dependent on the number of Pelmanism rounds included in the training session. In my study, participants completed three rounds whereas participants in Crundall et al.'s (2017) study completed five rounds. Although a short training intervention may be beneficial for maintaining drivers' motivation, it is likely that the motorcycle Pelmanism training task needed more than three rounds for drivers to demonstrate any observable training benefits. The balance between the length of the training and drivers' motivation levels is thus crucial for the motorcycle Pelmanism task to become a successful training tool and needs to be further researched, which will be discussed later. Given that this is the first piece of contradictory evidence and the second empirical study examining the effectiveness of a Pelmanism-based motorcycle perceptual training task, there is currently insufficient evidence to disregard the potential of a Pelmanism-based training task. In addition, there is some evidence that the Pelmanism matching task may have a positive effect on improving perceptual abilities, although the effect could not be attributed to the motorcycle-specific Pelmanism training task per se. Despite being exposed to more car trials than motorcycle trials in the T-junction task, participants in both training groups (i.e., motorcycle Pelmanism and pedestrian Pelmanism tasks) in Experiment 8 improved in motorcycle detection to a greater extent than car detection. This could suggest that there is generic training effect that is not solely due to practice effects. As drivers have a natural tendency to process wide objects before narrow objects (Navon, 1997; Rogé et al., 2010), a subordinate-level Pelmanism training task may provide a secondary benefit that mitigates the tendency to overlook high spatial frequency objects. In other words, by matching pairs of stimuli at a subordinate level, participants may have been primed to process information at a high level rather than a global level. In line with this, the priming effect was not present when participants were tested again after a delay of four weeks. However, empirical evidence is needed to demonstrate the causality between Pelmanism conducted at a subordinate level and improved motorcycle perception. Future directions for such research will be discussed later in the next section.

Compared to Crundall et al.'s (2017) version of the motorcycle Pelmanism training task, several improvements were made that contributed to the development of a motorcycle-specific training tool. First, the use of motorcycle stimuli with riders on road backgrounds ensures that participants are exposed to naturalistic stimuli during the training. This was an improvement to Crundall et al.'s (2017) version, which used images of motorcycles without riders, because the presence of a rider changes the shape or outline of a motorcycle hazard. Furthermore, the previous version used stimuli with a plain white background whereas the current version used roads as backgrounds to the motorcycle stimuli. The current version of stimuli more closely resembles real life situations when drivers are scanning the road for hazards as compared to a white background. Another improvement to the motorcycle Pelmanism training task is the addition of game elements that help motivate and engage drivers in order to address the lack of incentives of voluntary driver training interventions. The game elements used in the motorcycle Pelmanism task were able to challenge and reward players appropriately for accomplishing goals. Therefore, the experiments presented in this thesis help advance the development of a motorcycle Pelmanism-based training task that is suitable to be implemented in the real world.

8.4.2 Implications and future directions for research

If Pelmanism is indeed capable of delivering perceptual training, it offers an engaging and relatively cost-efficient approach to improving drivers' motorcycle perceptual skills. Current research on the motorcycle Pelmanism-based training task is only at an elementary stage, and continued research is needed before ruling out the use of the motorcycle Pelmanism task as a perceptual training tool. A potential direction for future research is to investigate why a positive training effect was reported by Crundall et al. (2017) but not in this thesis. For instance, future research could explore how differences in features of the training could impact the effectiveness of the motorcycle Pelmanism training task. In particular, studies could be conducted to investigate whether a motorcycle Pelmanism task with five rounds produces a training effect as opposed to a motorcycle Pelmanism task with three rounds, and whether motivation levels would be adversely affected by the increase in number of rounds.

Furthermore, research is needed to understand why a general training effect was found across both motorcycle and pedestrian Pelmanism conditions in my final experiment. As suggested earlier, perhaps training drivers to process information at a high-level of detail lowers drivers' tendency to favour wide objects over narrow objects, thus reducing the likelihood that they overlook motorcycles on the road. Based on this reasoning, an experiment could be conducted to compare the effectiveness of a Pelmanism-based training task using high-spatial frequency stimuli, such as motorcycles, and a Pelmanism-based training task using low-spatial frequency stimuli, such as cars or buses. If a training effect is observed in the high-spatial frequency condition but not the low-spatial frequency condition, this would support the hypothesis that the matching task primed drivers to process high-spatial frequency objects. Furthermore, if Pelmanism training with high-spatial frequency stimuli is found to be capable of improving motorcycle detection rates, studies could explore how to prolong the priming effect that ostensibly inhibited the bias against narrow objects in the visual processing hierarchy. There is currently insufficient information to determine how long this training or priming effect would last. This could offer another perspective to current efforts in the development of a motorcycle-specific perceptual training tool.

8.5 Conclusion

The aim of this thesis was twofold – to investigate whether dual drivers have an advantage at motorcycle perception and to evaluate the effectiveness of a Pelmanism-based training task for improving car drivers' motorcycle perception. Several conclusions can be drawn based on the findings of this thesis. The first and third conclusion directly address the two main research aims, and the second conclusion was drawn based on findings that contributed to the second research aim.

1) The first conclusion is that evidence for a dual driver perceptual advantage for motorcycles is weak – out of four studies comparing dual drivers against car drivers, the only task that found a motorcycle-specific advantage for dual drivers was a memory matching task. The current findings provide a strong rationale for researchers to revisit the assumption that drivers with motorcycle riding experience have a

perceptual advantage for spotting other motorcyclists on the road. Furthermore, they highlight the need to establish a causal relationship between dual drivers' collision rates and perceptual skills, using an updated sample of collision statistics.

2) Secondly, game elements can be used in driver training interventions to enhance motivation and willingness to take part. The motorcycle Pelmanism training task, specifically, benefits from an addition of time-based goals, points and competition. Accuracy-based goals were less effective than time-based goals, likely because motivation to be accurate occurs in short bursts whereas the sense of urgency arising from the motivation to be quick heightens throughout the game.

3) The third conclusion is that the particular version of the motorcycle Pelmanism task developed in this thesis was not effective in improving drivers' motorcycle perception measured by the three perceptual tasks. In this version, several notable differences, including the number of rounds and the use of a control pedestrian Pelmanism task with high-level of details, could have resulted in a different conclusion drawn from previous evidence. The lack of training effects reported in this thesis should not thwart subsequent research into a Pelmanism-based training intervention. With further research and improvements, an effective motorcycle Pelmanism-based training task may be possible. The prevalence of motorcycles involved in collisions caused by other drivers is an incentive to continue efforts into improving motorcycle safety by reducing such collisions.

References

Abdi, H., & Williams, L. J. (2010). Contrast analysis. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 243–251). http://dx.doi.org/10.4135/9781412961288

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. *Personality and Social Psychology Bulletin*, *38*(3), 317–330. https://doi.org/10.1177/0146167211427147

ACEM. (2009). *MAIDS Final Report 3rd Draft*. Retrieved from http://www.maids-study.eu/pdf/MAIDS2.pdf

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alberti, C. F., Shahar, A., & Crundall, D. (2014). Are experienced drivers more likely than novice drivers to benefit from driving simulations with a wide field of view? *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(Part A), 124–132. https://doi.org/10.1016/j.trf.2014.09.011

Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science*, *10*(3), 249–255. https://doi.org/10.1111/1467-9280.00145

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Macmillan Education UK

Bandura, A. (1982). Self-Efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of mental health* (Vol. 4, pp. 71-81). Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Beanland, V., Goode, N., Salmon, P. M., & Lenné, M. G. (2013). Is there a case for driver training? A review of the efficacy of pre- and post-licence driver training. *Safety Science*, *51*(1), 127–137. https://doi.org/10.1016/j.ssci.2012.06.021

Beanland, V., Lenne, M. G., & Underwood, G. (2015). Can drivers' expectations and behaviour around motorcycles be influenced by exposure? In L. Rößger, M. G. Lenné, & G. Underwood (Eds.), *Increasing motorcycle conspicuity: Design and assessment of interventions to enhance rider safety* (pp. 165–182). Ashgate Publishing.

Bechhofer, R. E., & Dunnett, C. W. (1982). Multiple comparisons for orthogonal contrasts: Examples and tables. *Technometrics*, *24*(3), 213–222. https://doi.org/10.1080/00401706.1982.10487761

Bentin, S., Moscovitch, M., & Nirhod, O. (1998). Levels of processing and selective attention effects on encoding in memory. *Acta Psychologica*, *98*(2–3), 311–341. https://doi.org/10.1016/s0001-6918(97)00048-6

Bonasia, K., Sekeres, M. J., Gilboa, A., Grady, C. L., Winocur, G., & Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. *Neurobiology of Learning and Memory*, *153*, 26–39.

https://doi.org/10.1016/j.nlm.2018.02.017

Borowsky, A., & Oron-Gilad, T. (2013). Exploring the effects of driving experience on hazard awareness and risk perception via real-time hazard identification, hazard classification, and rating tasks. *Accident Analysis and Prevention*, *59*, 548–565. https://doi.org/10.1016/j.aap.2013.07.008

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention*, *42*(4), 1240–1249. https://doi.org/10.1016/j.aap.2010.02.001

Bostan, B., & Öğüt, S. (2009). Game challenges and difficulty fevels: Lessons learned from RPGs. *International simulation and gaming association conference.*

Brown, I. D. (2002). *A review of the looked but failed to see accident causation factor*.

Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Attention, Perception, and Psychophysics*, *72*(7), 1865–1874. https://doi.org/10.3758/APP.72.7.1865

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Cameron, J., Pierce, W. D., Banko, K. M., & Gear, A. (2005). Achievement-based rewards and intrinsic motivation: A test of cognitive mediators. *Journal of Educational Psychology*, *97*(4), 641–655. https://doi.org/10.1037/0022-0663.97.4.641

Cantwell, S. J., Isler, R. B., & Starkey, N. J. (2013). The effects of road commentary training on novice drivers' visual search behaviour: A preliminary investigation. *Proceedings of the 2013 Australasian Road Safety Research, Policing & Education Conference, Brisbane, Queensland.*

Castro, C., Ventsislavova, P., Peña-Suarez, E., Gugliotta, A., Garcia-Fernandez, P., Eisman, E., & Crundall, D. (2016). Proactive Listening to a Training Commentary improves hazard prediction. *Safety Science*, *82*, 144–154. https://doi.org/10.1016/j.ssci.2015.09.018

Causer, J., Barach, P., & Williams, A. M. (2014). Expertise in medicine: using the expert performance approach to improve simulation training. *Medical Education*, *48*(2), 115–123. https://doi.org/10.1111/medu.12306

Cavallo, V., & Pinto, M. (2012). Are car daytime running lights detrimental to motorcycle conspicuity? *Accident Analysis and Prevention*, *49*, 78–85. https://doi.org/10.1016/j.aap.2011.09.013

Cavallo, V., Ranchet, M., Pinto, M., Espié, S., Vienne, F., & Dang, N. T. (2015). Improving car drivers' perception of motorcycle motion through innovative headlight configurations. *Accident Analysis and Prevention*, *81*, 187–193. https://doi.org/10.1016/j.aap.2015.04.034

Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 157–167. https://doi.org/10.1016/S1369-8478(02)00014-1

Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. *Games and Culture*, *5*(2), 177–198. https://doi.org/10.1177/1555412009354727

Chase, W. G., & Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology*, *4,* 55-81.

Cheng, H. N. H., Wu, W. M. C., Liao, C. C. Y., & Chan, T. W. (2009). Equal opportunity tactic: Redesigning and applying competition games in classrooms. *Computers and Education*, *53*(3), 866–876. https://doi.org/10.1016/j.compedu.2009.05.006

Cheong, C., Filippou, J., & Cheong, F. (2014). Towards the gamification of learning: Investigating student perceptions of game elements. *Journal of Information Systems Education, 25*(3), 233-244.

Christensen, R. H. B. (2019). "*ordinal: Regression models for ordinal data [R package ordinal version 2019.12-10]*. Retrieved from https://cran.r-project.org/package=ordinal

Christie, J., & Klein, R. (1995). Familiarity and attention: Does what we know affect what we notice? *Memory & Cognition, 23*(5), 547-550. https://doi.org/10.3758/BF03197256

Clarke, D. D., Ward, P., Bartle, C., & Truman, W. (2004). *In-depth study of motorcycle accidents* (Road Safety Research Report No. 54). Retrieved from Department for Transport: http://speedcamerareport.co.uk/dft_motorcycle_accidents.pdf

Clarke, D. D., Ward, P., Bartle, C., & Truman, W. (2007). The role of motorcyclist and other driver behaviour in two types of serious accident in the UK. *Accident Analysis and Prevention*, *39*(5), 974–981. https://doi.org/10.1016/j.aap.2007.01.002

Comelli, M., Morandi, A., Magazzù, D., Bottazzi, M., & Marinoni, A. (2008). Brightly coloured motorcycles and brightly coloured motorcycle helmets reduce the odds of a specific category of road accidents: A case-control study. *Biomedical Statistics and Clinical Epidemiology*, *2*(1), 71–78.

Cooper K. M., Brownell S. E. (2020) Student anxiety and fear of negative evaluation in active learning science classrooms. In J. J. Mintzes & E. M. Walter (Eds.) *Active Learning in College Science* (pp.909-925). Springer, Cham. https://doi.org/10.1007/978-3-030-33600-4_56

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis and Prevention*, *86*, 47–58. https://doi.org/10.1016/j.aap.2015.10.006

Crundall, D., Andrews, B., Van Loon, E., & Chapman, P. (2010). Commentary training improves responsiveness to hazards in a driving simulator. *Accident Analysis and Prevention*, *42*(6), 2117–2124. https://doi.org/10.1016/j.aap.2010.07.001

Crundall, D., Clarke, D., Ward, P., & Bartle, C. (2008). *Car drivers' skills and attitudes to motorcycle safety: A review* (Road Safety Research Report No. 85). Retrieved from Department for Transport: http://www.righttoride.co.uk/virtuallibrary/ridersafety/DfTStudycardriverattitudetowardsmotorcyclists.pdf

Crundall, D., Crundall, E., Clarke, D., & Shahar, A. (2012). Why do car drivers fail to give way to motorcycles at t-junctions? *Accident Analysis and Prevention*, *44*(1), 88–96. https://doi.org/10.1016/j.aap.2010.10.017

Crundall, D., Howard, A., & Young, A. (2017). Perceptual training to increase drivers' ability to spot motorcycles at T-junctions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *48*, 1–12. https://doi.org/10.1016/j.trf.2017.05.003

Crundall, D., Humphrey, K., & Clarke, D. (20M.08). Perception and appraisal of approaching motorcycles at junctions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(3), 159–167. https://doi.org/10.1016/j.trf.2007.09.003

Crundall, E., Stedmon, A. W., Saikayasit, R., & Crundall, D. (2013). A simulator study investigating how motorcyclists approach side-road hazards. *Accident Analysis and*

Prevention, *51*, 42–50. https://doi.org/10.1016/j.aap.2012.10.017

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row.

Curby, K. M., & Gauthier, I. (2009). The temporal advantage for individuating objects of expertise: Perceptual expertise is an early riser. *Journal of Vision*, *9*(6), 7–7. https://doi.org/10.1167/9.6.7

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A Visual Short-Term Memory Advantage for Objects of Expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94–107. https://doi.org/10.1037/0096-1523.35.1.94

Davidson, D. J., & Martin, A. E. (2013). Modeling accuracy as a function of response time with the generalized linear mixed effects model. *Acta Psychologica*, *144*(1), 83–96. https://doi.org/10.1016/j.actpsy.2013.04.016

de la Rosa, S., Choudhery, R. N., & Chatziastros, A. (2010). Visual object detection, categorization, and identification tasks are associated with different time courses and sensitivities. *Journal of Experimental Psychology: Human Perception & Performance, 37*(1), 38-47. https://doi.org/10.1037/a0020553

DeLucia, P. R. (1991). Pictorial and motion-based information for depth perception. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 738–748. https://doi.org/10.1037//0096-1523.17.3.738

Department for Transport. (2014). *The accidents sub-objective TAG unit 3.4.1.* Retrieved from https://webarchive.nationalarchives.gov.uk/20140304110017/http://www.dft.gov.uk/webtag/documents/expert/unit3.4.1.php

Department for Transport. (2015a). *Facts on motorcyclist casualties*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/447673/motorcyclist-casualties-2013-data.pdf

Department for Transport. (2015b). *Facts on pedal cyclists*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/447674/pedal-cyclists-2013-data.pdf

Department for Transport. (2019a). *Reported road casualties in Great Britain: 2018 annual report*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/834585/reported-road-casualties-annual-report-2018.pdf

Department for Transport. (2019b). Accident and casualty costs (RAS60001) [Data table]. Retrieved from https://www.gov.uk/government/statistical-data-sets/ras60-average-value-of-preventing-road-accidents

Department for Transport. (2019c). *Road traffic estimates in Great Britain: 2018 report*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/808555/road-traffic-estimates-in-great-britain-2018.pdf

Department for Transport. (2019d). Vehicles involved in reported road accidents (RAS20001) [Data table]. Retrieved from https://www.gov.uk/government/statistical-data-sets/ras20-drivers-riders-and-vehicles-in-reported-road-accidents

Department for Transport. (2019e). Vehicles involved in reported road accidents (RAS20008) [Data table]. Retrieved from https://www.gov.uk/government/statistical-data-sets/ras20-drivers-riders-and-vehicles-in-reported-road-accidents

Department for Transport. (2020a). *Story of THINK!* https://www.think.gov.uk/about-think/story-of-think/

Department for Transport. (2020b). Take longer to look [Image]. Retrieved from https://www.think.gov.uk/campaign/motorcycling/#group-posters-2

Desurvire, H., Caplan, M., & Toth, J. A. (2004). Using heuristics to evaluate the playability of games. *In CHI'04 Extended Abstracts on Human Factors in Computing Systems,* 1509–1512. https://doi.org/10.1145/985921.986102

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments,* 9-15. https://doi.org/10.1145/2181037.2181040

England, B. J., Brigati, J. R., & Schussler, E. E. (2017). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PLOS ONE*, *12*(8), e0182506. https://doi.org/10.1371/journal.pone.0182506

Ennever, B. (2020, June 2). *The Pelman School of Memory, The Pelman Institute and Pelmanism.* Ennever. http://www.ennever.com/histories/history386p.php

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits.* Cambridge University Press.

Field, A. P., & Wright, D. B. (2011). A primer on using multilevel models in clinical and experimental psychopathology research. *Journal of Experimental Psychopathology*, *2*(2), 271–293. https://doi.org/10.5127/jep.013711

Fisher, D. L., Pollatsek, A. P., & Pradhan, A. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash?. *Injury Prevention*, *12*(suppl 1), i25-i29. https://doi.org/10.1136/ip.2006.012021

Fülöp, M. (2009). Happy and Unhappy Competitors: What Makes the Difference? *Psychological Topics, 18*(2), 345-367.

Galpin, A., Underwood, G., & Crundall, D. (2009). Change blindness in driving scenes.

Transportation Research Part F: Traffic Psychology and Behaviour, *12*(2), 179–185. https://doi.org/10.1016/j.trf.2008.11.002

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441–467. https://doi.org/10.1177/1046878102238607

Gayet, S., van Maanen, L., Heilbron, M., Paffen, C. L. E., & Van der Stigchel, S. (2016). Visual input that matches the content of visual working memory requires less (not faster) evidence sampling to reach conscious access. *Journal of Vision*, *16*(11), 26–26. https://doi.org/10.1167/16.11.26

Gershon, P., Ben-Asher, N., & Shinar, D. (2012). Attention and search conspicuity of motorcycles as a function of their visual context. *Accident Analysis and Prevention*, *44*(1), 97–103. https://doi.org/10.1016/j.aap.2010.12.015

Gobet, F. (2005). Chunking models of expertise: implications for education. *Applied Cognitive Psychology*, *19*(2), 183–204. https://doi.org/10.1002/acp.1110

Godwin, H. J., Menneer, T., Riggs, C. A., Cave, K. R., & Donnelly, N. (2015). Perceptual failures in the selection and identification of low-prevalence targets in relative prevalence visual search. *Attention, Perception, and Psychophysics*, *77*(1), 150–159. https://doi.org/10.3758/s13414-014-0762-8

Golan, T., Bentin, S., Degutis, J. M., Robertson, L. C., & Harel, A. (2013). Association and dissociation between detection and discrimination of objects of expertise: Evidence from visual search. *Attention, Perception, & Psychophysics, 76,* 391-406. https://doi.org/10.3758/s13414-013-0562-6

Goodge, T., Kroll, V., Vernon, M., Ventsislavova, P., & Crundall, D. (submitted). Assessing appropriateness and cybersickness in a 360-degree hazard awareness test. *Applied Ergonomics.*

Gould, M., Poulter, D. R., Helman, S., & Wann, J. P. (2012a). Errors in judging the approach rate of motorcycles in nighttime conditions and the effect of an improved lighting configuration. *Accident Analysis and Prevention*, *45*, 432–437. https://doi.org/10.1016/j.aap.2011.08.012

Gould, M., Poulter, D. R., Helman, S., & Wann, J. P. (2012b). Judgments of approach speed for motorcycles across different lighting levels and the effect of an improved tri-headlight configuration. *Accident Analysis and Prevention*, *48*, 341–345. https://doi.org/10.1016/j.aap.2012.02.002

Graafland, M., Schraagen, J. M. C., Schijven, M. P. (2012). Systematic review of validity of serious games for medical education and surgical skills training. *The British Journal of Surgery*, *99*(10), 1322–1330.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–160. https://doi.org/10.1111/j.0956-7976.2005.00796.x

Grund, C. K. (2015). How games and game elements facilitate learning and motivation: A literature review. In: D. W. Cunningham, P. Hofstedt, K. Meer, & I. Schmitt (Eds.), *INFORMATIK 2015. Bonn: Gesellschaft für Informatik e.V.* (pp. 1279-1293).

Gunter, G. A., Robert F Kenny, P. D., & Erik Henry Vick, P. D. (2006). A case for a formal design paradigm for serious games. *The Journal of the International Digital Media and*

Arts Association, 3(1), 93-105.

Guznov, S., Matthews, G., Warm, J. S., & Pfahler, M. (2017). Training techniques for visual search in complex task environments. *Human Factors*, *59*(7), 1139–1152. https://doi.org/10.1177/0018720817712307

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Han, S., Yund, E. W., & Woods, D. L. (2003). An ERP study of the global precedence effect: The role of spatial frequency. *Clinical Neurophysiology*, *114*(10), 1850–1865. https://doi.org/10.1016/S1388-2457(03)00196-2

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers and Education*, *80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Harel, A. (2016). What is special about expertise? Visual expertise reveals the interactive nature of real-world object recognition. *Neuropsychologia*, *83*, 88–99. https://doi.org/10.1016/j.neuropsychologia.2015.06.004

Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cerebral Cortex*, *20*(10), 2304–2318. https://doi.org/10.1093/cercor/bhp316

Harel, A., Kravitz, D., & Baker, C. I. (2013). Beyond perceptual expertise: Revisiting the neural substrates of expert object recognition. *Frontiers in Human Neuroscience*, *7*(885). https://doi.org/10.3389/fnhum.2013.00885

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, *32*(4), 385–410. https://doi.org/10.1111/j.1468-2958.2006.00281.x

Hershler, O., & Hochstein, S. (2009). The importance of being expert: Top-down attentional control in visual search with photographs. *Attention, Perception, and Psychophysics*, *71*(7), 1478–1486. https://doi.org/10.3758/APP.71.7.1478

Herzmann, G., & Curran, T. (2011). Experts' memory: An ERP study of perceptual expertise effects on encoding and recognition. *Memory and Cognition*, *39*(3), 412–432. https://doi.org/10.3758/s13421-010-0036-1

Hole, G. J., Tyrrell, L., & Langham, M. (1996). Some factors affecting motorcyclists' conspicuity. *Ergonomics*, *39*(7), 946–965. https://doi.org/10.1080/00140139608964516

Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2),388-403. https://doi.org/10.1037/0096-1523.29.2.388

Hollingworth A. (2012). Guidance of visual search by memory and knowledge. In M. D. Dodd, J. H. Flowers (Eds.), *The Influence of Attention, Learning, and Motivation on Visual Search. Nebraska Symposium on Motivation.* Springer, New York, NY. https://doi.org/10.1007/978-1-4614-4794-8_4

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113-136. https://doi.org/10.1037/0096-1523.28.1.113

Horswill, M. S. (2016). Hazard perception in driving. *Current Directions in Psychological Science*, *25*(6), 425–430. https://doi.org/10.1177/0963721416663186

Horswill, M. S., & Helman, S. (2003). A behavioral comparison between motorcyclists and a matched group of non-motorcycling car drivers: Factors influencing accident risk.

Accident Analysis and Prevention, *35*(4), 589–597. https://doi.org/10.1016/S0001-4575(02)00039-8

Horswill, M. S., Helman, S., Ardiles, P., & Wann, J. P. (2005). Motorcycle accident risk could be inflated by a time to arrival illusion. *Optometry and Vision Science. 82*(8), 740-746. https://doi.org/10.1097/01.opx.0000175563.21423.50

Horswill, M. S., Kemala, C. N., Wetton, M., Scialfa, C. T., & Pachana, N. A. (2010). Improving older drivers' hazard perception ability. *Psychology and Aging*, *25*(2), 464–469. https://doi.org/10.1037/a0017306

Horswill, M. S., Marrington, S. A., Mccullough, C. M., Wood, J., Pachana, N. A., Mcwilliam, J., & Raikos, M. K. (2008). The hazard perception ability of older drivers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 63*(4), 212-218. https://doi.org/10.1093/geronb/63.4.P212

Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness* (pp.155-175). Aldershot, UK: Ashgate.

Horswill, M. S., Sullivan, K., Lurie-Beck, J. K., & Smith, S. (2013). How realistic are older drivers' ratings of their driving ability? *Accident Analysis and Prevention*, *50*, 130–137. https://doi.org/10.1016/j.aap.2012.04.001

Horswill, M. S., Taylor, K., Newnam, S., Wetton, M., & Hill, A. (2013). Even highly experienced drivers benefit from a brief hazard perception training intervention. *Accident Analysis and Prevention*, *52*, 100–110. https://doi.org/10.1016/j.aap.2012.12.014

Horswill, M. S., Waylen, A. E., & Tofield, M. I. (2004). Drivers' Ratings of Different Components of Their Own Driving Skill: A Greater Illusion of Superiority for Skills

That Relate to Accident Involvement. *Journal of Applied Social Psychology*, *34*(1), 177–195. https://doi.org/10.1111/j.1559-1816.2004.tb02543.x

Huang, B., & Preston, J. (2004). *A literature review on motorcycle collisions: Final report*. Retrieved from Transport Studies Unit, Oxford University: https://pdfs.semanticscholar.org/c0d9/0bf237e48c8ba7fbaeb1e8c60c93d4d651c3.pdf

Hurt, H. H., Ouellet, J. V., & Thom, D. R. (1981). *Motorcycle accident cause factors and identification of countermeasures.* Volume 1: Technical report (No. 81-1). National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/6450

Isler, R. B., Starkey, N. J., & Williamson, A. R. (2009). Video-based road commentary training improves hazard perception of young drivers in a dual task. *Accident Analysis and Prevention*, *41*(3), 445–452. https://doi.org/10.1016/j.aap.2008.12.016

Jackson, L., Chapman, P., & Crundall, D. (2009). What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics*, *52*(2), 154–164. https://doi.org/10.1080/00140130802030714

Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 556–568. https://doi.org/10.1037/0096-1523.34.3.556

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829-6833. https://doi.org/10.1073/pnas.0801268105

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. https://doi.org/10.1037/a0028347

Kalakoski, V., & Saariluoma, P. (2001). Taxi drivers' exceptional memory of street names. *Memory and Cognition*, *29*(4), 634–638. https://doi.org/10.3758/BF03200464

Karni, A., & Bertini, G. (1997). Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, *7*(4), 530–535. https://doi.org/10.1016/S0959-4388(97)80033-5

Keyes, H., Green, F., Compton, C., & Staton, M. (2019). Short-term cognitive conspicuity training does not improve driver detection of motorcycles at road junctions: A reply to Crundall, Howard, and Young (2017). *Transportation Research Part F: Traffic Psychology and Behaviour*, *67*, 57–65. https://doi.org/10.1016/j.trf.2019.10.008

Kinzer, C. K., Hwang, M., Chantes, P., Choi, A., & Hsu, S. (2015). *Design decisions and educational games: Insights for acceptance [Conference presentation]*. Games, Learning & Society Conference, Madison, Wisconsin, United States.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Krajicek, M. E., & Schears, R. M. (2010). Daytime running lights in the USA: What is the impact on vehicle crashes in Minnesota? *International Journal of Emergency Medicine*, *3*(1), 39–43. https://doi.org/10.1007/s12245-009-0151-6

Labbett, S., & Langham, M. (2006). What do drivers do at junctions? In *71st RoSPA Road Safety Congress*. Blackpool, England.

Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics*, *54*(6), 773–784. https://doi.org/10.3758/BF03211802

Landers, R. N., Armstrong, M. B., & Collmus, A. B. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. In M. Ma & A. Oikonomou (Eds.)*, Serious games and edutainment applications* (pp. 457-483). Springer, Cham. https://doi.org/10.1007/978-3-319-51645-5_21

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behaviour, 71,* 508-515. https://doi.org/10.1016/j.chb.2015.08.008

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*, *45*(6), 769–785. https://doi.org/10.1177/1046878114563662

Langham, M., Hole, G., Edwards, J., & O'Neil, C. (2002). An analysis of "looked but failed to see" accidents involving parked police vehicles. *Ergonomics*, *45*(3), 167–185. https://doi.org/10.1080/00140130110115363

Latham, G. P. (2004). The motivational benefits of goal-setting. *Academy of Management Executive*, *18*(4), 126-129.

Lee, Y. M., & Sheppard, E. (2017). Differences in gap acceptance for approaching cars and motorcycles at junctions: What causes the size-arrival effect? *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 50–54. https://doi.org/10.1016/j.trf.2017.07.006

Lee, Y. M., Sheppard, E., & Crundall, D. (2015). Cross-cultural effects on the perception and appraisal of approaching motorcycles at junctions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *31*, 77–86. https://doi.org/10.1016/j.trf.2015.03.013

Lieberoth, A. (2015). Shallow gamification-testing psychological effects of framing an activity as a game. *Games and Culture, 10*(3), 229-248. https://doi.org/10.1177/1555412014559978

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation a 35-year odyssey. *American Psychologist, 57*(9), 705-717. https://doi.org/10.1037/0003-066X.57.9.705

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious Games Analytics: Theoretical Framework. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*. Springer. https://doi.org/10.1007/978-3-319-05834-4_1

Ludwig, C. J. H., Gilchrist, I. D., McSorley, E., & Baddeley, R. J. (2005). The temporal impulse response underlying saccadic decisions. *Journal of Neuroscience*, *25*(43), 9907–9912. https://doi.org/10.1523/JNEUROSCI.2197-05.2005

Lunenburg, F. C. (2011). Goal-setting theory of motivation. *International Journal of Management, Business, and Administration, 15*(1), 1-6.

Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(4), 794–806. https://doi.org/10.1037/xhp0000343.supp

Lynam, D., Broughton, J., Minton, R., & Tunbridge, R. J. (2001). *An analysis of police reports of fatal accidents involving motorcycles* (TRL Report 492). Retrieved from http://trl.demo.varistha.co.uk/uploads/trl/documents/TRL492.pdf

Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5),1067-1079. https://doi.org/10.1037/a0020254

Mack, A., & Rock, I. (1998). Inattentional blindness: Perception without attention. In R. D. Wright (Ed.), *Visual attention* (pp. 55–76). Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology

Press.

Magazzù, D., Comelli, M., & Marinoni, A. (2006). Are car drivers holding a motorcycle licence less responsible for motorcycle-car crash occurrence?: A non-parametric approach. *Accident Analysis and Prevention*, *38*(2), 365–370. https://doi.org/10.1016/j.aap.2005.10.007

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10*(1), 69-75. https://doi.org/10.3758/BF03197627

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 2857–2866. https://doi.org/10.1145/1978942.1979366

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research*, *11*(1), 77–95. https://doi.org/10.1016/S0926-6410(00)00066-5

Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision*, *14*(12), 1–1. https://doi.org/10.1167/14.12.1

Meir, A., Borowsky, A., & Oron-Gilad, T. (2014). Formation and evaluation of Act and Anticipate Hazard Perception Training (AAHPT) intervention for young novice drivers. *Traffic Injury Prevention*, *15*(2), 172–180. https://doi.org/10.1080/15389588.2013.802775

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048

Ministry of Transport. (2020). *Motorcycle crash facts 2017.* Retrieved from https://www.transport.govt.nz/mot-resources/road-safety-resources/crashfacts/motorcyclecrashfacts/

Most, S. B., & Astur, R. S. (2007). Feature-based attentional set as a cause of traffic accidents. *Visual Cognition*, *15*(2), 125–132. https://doi.org/10.1080/13506280600959316

Motorcycle Action Group. (2006). *How close is too close? Concerning car collisions and motorcycles*. Retrieved from http://motorcycleminds.org/virtuallibrary/ridersafety/HowCloseisTooClose.pdf

Musselwhite, C. B. A., Avineri, E., Susilo, Y. O., & Bhattachary, D. (2012). Public attitudes towards motorcyclists' safety: A qualitative study from the United Kingdom. *Accident Analysis and Prevention*, *49*, 105–113. https://doi.org/10.1016/j.aap.2011.06.005

Mutter, T., & Kundisch, D. (2014). Behavioral mechanisms prompted by badges: The goal-gradient hypothesis. *ICIS*.

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford University Press.

Nako, R., Wu, R., & Eimer, M. (2014). Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 50–60. https://doi.org/10.1037/a0033228

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? *Journal of Experimental Psychology: Human Perception and Performance, 40*(4),1283-1288. https://doi.org/10.1037/a0036885

National Highway Traffic Safety Administration. (2019). *Traffic safety facts 2017 data: Motorcycles*. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812785

Navon, D. (1977). The precedence of global features in visual perception. *Cognitive Psychology*, *383*. https://doi.org/10.1179/095066068790421683

Olivers, C. N. L. (2011). Long-term visual associations affect attentional guidance. *Acta Psychologica*, *137*(2), 243–247. https://doi.org/10.1016/j.actpsy.2010.07.001

Pai, C. W. (2011). Motorcycle right-of-way accidents - A literature review. *Accident Analysis and Prevention*, *43*(3), 971–982. https://doi.org/10.1016/j.aap.2010.11.024

Pai, C. W., Hwang, K. P., & Saleh, W. (2009). A mixed logit analysis of motorists' right-of-way violation in motorcycle accidents at priority T-junctions. *Accident Analysis and Prevention*, *41*(3), 565–573. https://doi.org/10.1016/j.aap.2009.02.007

Pai, C. W., & Saleh, W. (2008). Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures. *Accident Analysis and Prevention*, *40*(2), 479–486. https://doi.org/10.1016/j.aap.2007.08.003

Paine, M., Paine, D., Haley, J., & Cockfield, S. (2005). Daytime running lights for motorcycles. In *19th International Technical Conference on the Enhanced Safety of Vehicles (ESV).* Washington DC, United States

Pammer, K., Sabadas, S., & Lentern, S. (2018). Allocating attention to detect motorcycles: The role of inattentional blindness. *Human Factors*, *60*(1), 5–19. https://doi.org/10.1177/0018720817733901

Peek-Asa, C., & Kraus, J. F. (1996). Injuries sustained by motorcycle riders in the approaching turn crash configuration. *Accident Analysis and Prevention*, *28*(5), 561–569. https://doi.org/10.1016/0001-4575(96)00022-X

Peltier, C., & Becker, M. W. (2016). Decision processes in visual search as a function of target prevalence. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1466-1476. https://doi.org/10.1037/xhp0000248

Pierce, W. D., & Cameron, J. (2002). A summary of the effects of reward contingencies on interest and performance. *The Behavior Analyst Today*, *3*(2), 221–228. https://doi.org/10.1037/h0099969

Pincus, D., Wasserstein, D., Nathens, A. B., Bai, Y. Q., Redelmeier, D. A., & Wodchis, W. P. (2017). Direct medical costs of motorcycle crashes in Ontario. *CMAJ*, *189*(46), E1410–E1415. https://doi.org/10.1503/cmaj.170337

Pinto, M., Cavallo, V., & Saint-Pierre, G. (2014). Influence of front light configuration on the visual conspicuity of motorcycles. *Accident Analysis and Prevention*, *62*, 230–237. https://doi.org/10.1016/j.aap.2013.09.026

Pradhan, A. K., & Crundall, D. (2016). Hazard avoidance in young novice drivers: Definitions and a framework. In D. L. Fisher, J. Caird, W. Horrey, & L. Trick (Eds.), *Handbook of teen and novice drivers: Research, practice, policy, and directions* (1st ed., pp. 61–74). Florida, US: CRC Press.

Pressley, A., Fernández-Medina, K., Helman, S., Mckenna, F. P., & Stradling, S. (2017). *A review of interventions which seek to increase the safety of young and novice drivers* (Published Project Report PPR781). Department for Transport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/609828/interventions-to-increase-young-and-novice-driver-safety.pdf

Psychology Software Tools. (2016). *E-prime 3.0.* Psychology Software Tools, Pittsburgh, USA. https://www.pstnet.com.

Public Health England. (2018). Retrieved from https://www.gov.uk/government/publications/health-profile-for-england-2018/chapter-

2-trends-in-mortality#leading-causes-of-death-in-2016

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rabe, M. M. (2018). *Generalized linear mixed modeling of signal detection theory* [Unpublished master's thesis]. University of Victoria.

Rawson, K. A., & Van Overschelde, J. P. (2008). How does knowledge promote memory? The distinctiveness theory of skilled memory. *Journal of Memory and Language*, *58*(3), 646–668. https://doi.org/10.1016/j.jml.2007.08.004

Radin, U. R., Mackay, M. G., & Hills, B. L. (1996). Modelling of conspicuity-related motorcycle accidents in Seremban and Shah Alam, Malaysia. *Accident Analysis and Prevention*, *28*(3), 325–332. https://doi.org/10.1016/0001-4575(95)00071-2

Reeder, R. R. (2017). Individual differences shape the content of visual representations. *Vision Research*, *141*, 266–281. https://doi.org/10.1016/j.visres.2016.08.008

Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *Journal of Vision*, *13*(3), 13. https://doi.org/10.1167/13.3.13

Reeder, R. R., Stein, T., & Peelen, M. V. (2016). Perceptual expertise improves category detection in natural scenes. *Psychonomic Bulletin and Review*, *23*(1), 172–179. https://doi.org/10.3758/s13423-015-0872-x

Reeder, R. R., van Zoest, W., & Peelen, M. V. (2015). Involuntary attentional capture by task-irrelevant objects that match the search template for category detection in natural scenes. *Attention, Perception, and Psychophysics*, *77*(4), 1070–1080.

https://doi.org/10.3758/s13414-015-0867-8

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–526. https://doi.org/10.1017/s0140525x03000104

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rice, A., Jonathon Phillips, P., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science, 24*(11), 2235-2243. https://doi.org/10.1177/0956797613492986

Richter, G., Raban, D. R., & Rafaeli, S. (2015). Studying gamification: The effect of rewards and incentives on motivation. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 21–46). Springer, Cham. https://doi.org/10.1007/978-3-319-10208-5_2

Road Safety Scotland [Scottish Government]. (2017, June 7). *VR road safety Scotland* [Video]. YouTube. https://www.youtube.com/watch?v=hnWgEGVjlak

Robbins, C. J., Allen, H. A., & Chapman, P. (2018). Comparing drivers' gap acceptance for cars and motorcycles at junctions using an adaptive staircase methodology. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 944–954. https://doi.org/10.1016/j.trf.2018.07.023

Robbins, C. J., Allen, H. A., Miller, K. A., & Chapman, P. (2019). The 'saw but forgot' error: A role for short-term memory failures in understanding junction crashes? *PLOS ONE*, *14*(9), e0222905. https://doi.org/10.1371/journal.pone.0222905

Robbins, A., & Hout, M. C. (2015). Categorical target templates: Typical category members are found and identified quickly during word-cued search. *Visual Cognition, 23*(7), 817-821. https://doi.org/10.1080/13506285.2015.1093247

Rogé, J., Douissembekov, E., & Vienne, F. (2012). Low conspicuity of motorcycles for car drivers: Dominant role of bottom-up control of visual attention or deficit of top-down control? *Human Factors*, *54*(1), 14–25. https://doi.org/10.1177/0018720811427033

Rogé, J., Ferretti, J., & Devreux, G. (2010). Sensory conspicuity of powered two-wheelers during filtering manoeuvres. *Le Travail Humain, 73*(1), 7-30.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22,* 511–524. https://doi.org/10.3758/BF03198390

Rosenbloom, T., Perlman, A., & Pereg, A. (2011). Hazard perception of motorcyclists and car drivers. *Accident Analysis and Prevention, 43*(3), 601-604. https://doi.org/10.1016/j.aap.2010.08.005

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 347–363. https://doi.org/10.1007/s11031-006-9051-8

Saenz, M., Buracas, G. T., & Boynton, G. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631-632. https://doi.org/10.1038/nn876

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis and Prevention*, *38*(2), 407–414. https://doi.org/10.1016/j.aap.2005.10.014

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, *51*(13), 1552–1566.

https://doi.org/10.1016/j.visres.2010.10.019

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers and Education*, *69*, 452–462. https://doi.org/10.1016/j.compedu.2013.07.004

Sanocki, T., Islam, M., Doyon, J. K., & Lee, C. (2015). Rapid scene perception with tragic consequences: observers miss perceiving vulnerable road users, especially in crowded traffic scenes. *Attention, Perception, and Psychophysics*, *77*(4), 1252–1262. https://doi.org/10.3758/s13414-015-0850-4

Santhanam, R., Liu, D., & Shen, W. C. M. (2016). Gamification of technology-mediated training: Not all competitions are the same. *Information Systems Research*, *27*(2), 453–465. https://doi.org/10.1287/isre.2016.0630

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, *11*, 53–60. https://doi.org/10.1038/nrn2737

Schmidt, J., & Zelinsky, G. J. (2009). Short article: Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914. https://doi.org/10.1080/17470210902853530

Schwark, J. D., MacDonald, J., Sandry, J., & Dolgov, I. (2013). Prevalence-based decisions undermine visual search. *Visual Cognition*, *21*(5), 541–568. https://doi.org/10.1080/13506285.2013.811135

Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S., & Wetton, M. (2011). A hazard perception test for novice drivers. *Accident Analysis and Prevention*, *43*(1), 204–208. https://doi.org/10.1016/j.aap.2010.08.010

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin and Review*, *15*(1), 215–222. https://doi.org/10.3758/PBR.15.1.21

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience*, *18*(9), 1453–1465. https://doi.org/10.1162/jocn.2006.18.9.1453

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *Journal of Human Computer Studies*, *74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–312. https://doi.org/10.1016/j.neuron.2007.06.015

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2004.11.006

Smeddinck, J. D., Mandryk, R. L., Birk, M. V, Gerling, K. M., Barsilowski, D., & Malaka, R. (2016). *How to present game difficulty choices? Exploring the impact on player experience*. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5595–5607). https://doi.org/10.1145/2858036.2858574

Stein, T., Reeder, R. R., & Peelen, M. V. (2016). Privileged access to awareness for faces and objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance, 42*(6), 788–798. https://doi.org/10.1037/xhp0000188

Summala, H., Pasanen, E., Räsänen, M., & Sievänen, J. (1996). Bicycle accidents and drivers' visual search at left and right turns. *Accident Analysis and Prevention*, *28*(2), 147–153. https://doi.org/10.1016/0001-4575(95)00041-0

Swan, G., Shahin, M., Albert, J., Herrmann, J., & Bowers, A. R. (2019). The effects of simulated acuity and contrast sensitivity impairments on detection of pedestrian hazards in a driving simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 213-226.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment*, *3*(3), 1–23. https://doi.org/10.1145/1077246.1077253

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*(2), 145–151. https://doi.org/10.1111/j.0956-7976.2005.00795.x

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457-482.

TNS BMRB. (2014). *THINK! Motorcycle safety campaign 2014 evaluation* [pdf]. Department for Transport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/417338/1._THINK__Motorbikes_2014.pdf

Transport for London. (2015). *Collisions and casualties on London's roads: Annual report 2015*. Retrieved from http://content.tfl.gov.uk/collisions-and-casulaties-on-londons-roads-annual-report-2015.pdf

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E., Stansifer, R. L., & Castellan, N. J. (1979). *Tri-level study of the causes of traffic accidents: Final report – Executive summary.* Institute for Research in Public Safety Indiana University. https://deepblue.lib.umich.edu/handle/2027.42/64993

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*(2), 58–64. https://doi.org/10.1016/j.tics.2006.11.009

Underwood, G., Crundall, D., & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(6), 435–446. https://doi.org/10.1016/j.trf.2011.04.008

Underwood, G., Humphrey, K., & van Loon, E. (2011). Decisions about objects in real-world scenes are influenced by visual saliency before and during their inspection. *Vision Research*, *51*(18), 2031–2038. https://doi.org/10.1016/j.visres.2011.07.020

Underwood, G., Ngai, A., & Underwood, J. (2013). Driving experience and situation awareness in hazard detection. *Safety Science*, *56*, 29–35. https://doi.org/10.1016/j.ssci.2012.05.025

van der Linden, M., Wegman, J., & Fernández, G. (2014). Task- and experience-dependent cortical selectivity to features informative for categorization. *Journal of Cognitive Neuroscience*, *26*(2), 319–333. https://doi.org/10.1162/jocn_a_00484

Van Elslande, P., & Faucher-Alberton, L. (1997). When expectancies become certainties: a potential adverse effect of experience. In T. Rothengatter, E. Carbonell Vaya (Eds.), *Traffic and transport psychology: Theory and application* (147-159). Elsevier.

van Roy, R., & Zaman, B. (2017). Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory. In Ma, M., & Oikonomou, A. (Eds.), *Serious Games and Edutainment Applications Volume II* (pp. 485–509). Springer International Publishing. https://doi.org/10.1007/978-3-319-51645-5_22

Vandercruysse, S., Vandewaetere, M., Leuven, K. U., Cornillie, F., & Clarebout, G. (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research and Development, 61*(6), 927-950. https://doi.org/10.1007/s11423-013-9314-5

Ventsislavova, P., & Crundall, D. (2018). The hazard prediction test: A comparison of free-response and multiple-choice formats. *Safety Science*, *109*, 246–255. https://doi.org/10.1016/j.ssci.2018.06.004

Ventsislavova, P., Gugliotta, A., Penã-Suarez, E., Garcia-Fernandez, P., Eisman, E., Crundall, D., & Castro, C. (2016). What happens when drivers face hazards on the road? *Accident Analysis and Prevention*, *91*, 43–54. https://doi.org/10.1016/j.aap.2016.02.013

Visordown News (2014). *Think Bike poster was edited 'to get message across.'* https://www.visordown.com/news/general/think-bike-poster-was-edited-'-get-message-across'

Vlahogianni, E. I., Yannis, G., & Golias, J. C. (2012). Overview of critical risk factors in power-two-wheeler safety. *Accident Analysis and Prevention*, *49*, 12–22. https://doi.org/10.1016/j.aap.2012.04.009

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*(1), 91–100. https://doi.org/10.1068/p5262

Wallis, T. S. A., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis and Prevention*, *39*(6), 1177–1185. https://doi.org/10.1016/j.aap.2007.03.003

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology, 33*(4), 448-457. https://doi.org/10.1037/h0027806

Waylen, A. E., Horswill, M. S., Alexander, J. L., & McKenna, F. P. (2004). Do expert drivers have a reduced illusion of superiority? *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*(4–5), 323–331. https://doi.org/10.1016/j.trf.2004.09.009

Wetton, M. A., Hill, A., & Horswill, M. S. (2013). Are what happens next exercises and self-generated commentaries useful additions to hazard perception training for novice drivers? *Accident Analysis and Prevention*, *54*, 57–66. https://doi.org/10.1016/j.aap.2013.02.013

Wells, S., Mullin, B., Norton, R., Langley, J., Connor, J., Lay-Yee, R., & Jackson, R. (2004). Motorcycle rider conspicuity and crash related injury: Case-control study. *British Medical Journal*, *328*(7444), 857–860. https://doi.org/10.1136/bmj.37984.574757.ee

Wells, P., Tong, S., Sexton, B., Grayson, G., & Jones, E. (2008). *Cohort II: A Study of Learner and New Drivers Volume 1-Main Report* (Road Safety Research Report No. 81). Retrieved from Department for Transport: https://www.broughtonschoolofmotoring.com/wp-content/uploads/cohrtiimainreport.pdf

Williams, M. J., & Hoffmann, E. R. (1979). Motorcycle conspicuity and traffic accidents. *Accident Analysis & Prevention, 11*(3), 209-224. https://doi.org/10.1016/0001-4575(79)90005-8

Wilson, S., Darling, S., & Sykes, J. (2011). Adaptive memory: Fitness relevant stimuli show a memory advantage in a game of pelmanism. *Psychonomic Bulletin and Review*, *18*(4), 781–786. https://doi.org/10.3758/s13423-011-0102-0

Wulf, G., Hancock, P. A., & Rahimi, M. (1989). Motorcycle conspicuity: An evaluation and synthesis of influential factors. *Journal of Safety Research*, *20*(4), 153–176. https://doi.org/10.1016/0022-4375(89)90025-X

World Health Organisation. (2018). Global status report on road safety 2018. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202-238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 11-55). Oxford University Press.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(0058). https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: Rare items often missed in visual searches. *Nature*, *435*(7041), 439–440. https://doi.org/10.1038/435439a

Wolfe, J. M., Horowitz, T. S., Wert, M. J. Van, Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136*(4), 623–638. https://doi.org/10.1037/0096-3445.136.4.623

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121–124. https://doi.org/10.1016/j.cub.2009.11.066

Wu, R., Pruitt, Z., Zinszer, B. D., & Cheung, O. S. (2017). Increased experience amplifies the activation of task-irrelevant category representations. *Attention, Perception, and Psychophysics*, *79*(2), 522–532. https://doi.org/10.3758/s13414-016-1254-9

Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory and Cognition*, *45*(4), 677–689. https://doi.org/10.3758/s13421-016-0679-7

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095–2103. https://doi.org/10.1016/j.visres.2009.05.017

Young, A. H., Chapman, P., & Crundall, D. (2014). Producing a commentary slows concurrent hazard perception responses. *Journal of Experimental Psychology: Applied*, *20*(3), 285–294. https://doi.org/10.1037/xap0000016

Young, A. H., Crundall, D., & Chapman, P. (2017). Commentary driver training: Effects of commentary exposure, practice and production on hazard perception and eye movements. *Accident Analysis and Prevention*, *101*, 1–10. https://doi.org/10.1016/j.aap.2017.01.007

Yuan, W. (2000). The effectiveness of the "ride-bright" legislation for motorcycles in Singapore. *Accident Analysis and Prevention*, *32*(4), 559–563. https://doi.org/10.1016/S0001-4575(99)00078-0

Zeidner, M., & Matthews, G. (2005). Evaluation Anxiety. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 141–163).

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps.* O'Reilly Media.

Appendix A

Description of clips used in the VR hazard prediction task Experiments 1 and 8

CLIP	DESCRIPTION
1	Driving along a high street, and oncoming car attempts to turn across your path into the road on the left, but cannot due to the no entry signs.
2	Travelling along the middle lane of an arterial road, the car ahead attempts to merge into the left lane, causing the car on the left to swerve in front of you in frustration.
3	Approaching a busy junction with an oncoming HGV, you intend to turn right. As you do, a motorcycle appears from behind the HGV and crosses your path.
4	Driving down a residential street with parked cars, an oncoming car flashes it's headlights to give way. A blue car appears from a side road in the left and blocks your path.
5	Travelling along a wide residential street with parked cars and pedestrians, a woman with a buggy steps out from behind a parked vehicle on the left.
6	Whilst driving along a suburban road approaching a junction. As you begin to turn left, a cyclist from the cycle lane on the left enters the road, and blocks your path
7	Travelling along a residential street, a pedestrian steps out from behind a parked car on the left.
8	As you overtake a bus waiting at a bus stop, a car appears from a side road on the left out of view, and blocks your path
9	Whilst driving down an empty suburban road, a pedestrian enters the zebra crossing ahead, hidden behind an oncoming car
10	As you are travelling along a suburban road, with heavy traffic in the oncoming lane, a motorcycle appears and blocks your path as you attempt to turn right.

Appendix B

Visual search arrays used in Experiment 2

Carphone
British Heart
Buchan Big Cat
GOODWIN


TESCO

Appendix C

Analysis without outliers for Experiment 2

Driver group was not a significant factor (χ^2 (1) = 0.11, p = .74), but time taken to find the target differed significantly according to target type (χ^2 (1) = 27.6, p < .001). Across both driver groups, participants took significantly longer to search for taxis (M = 9.36s, SD = 9.72) than motorcycles (M = 3.99s, SD = 3.89). The interaction between driver group and target was not significant (χ^2 (1) = 0.64, p = .42; see Figure C1 for means).

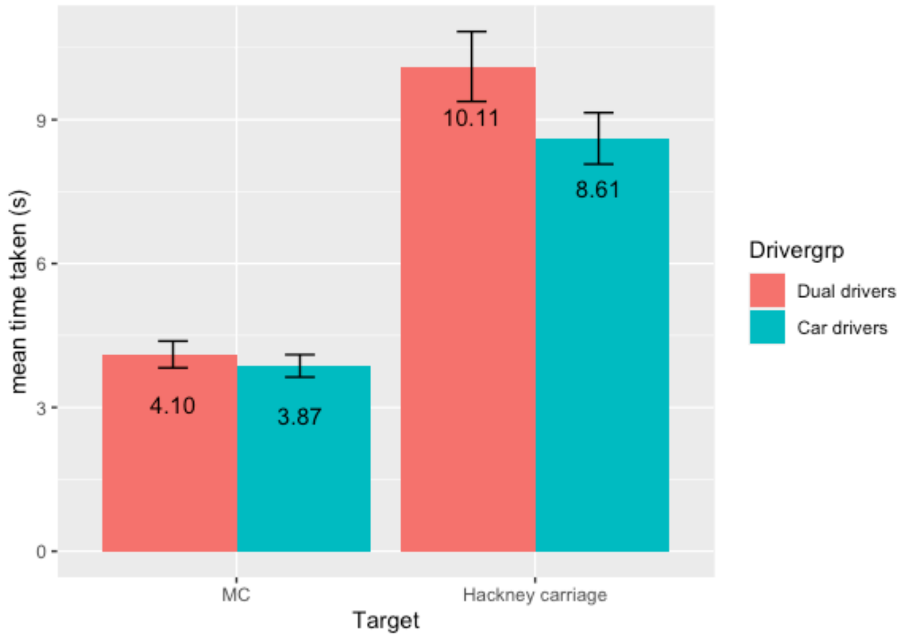

Figure C1: Mean time taken with +1/-1 SE to locate motorcycles or hackney carriages by driver group (without outliers)

Appendix D

Model fit comparison for Experiment 4

Source	AIC	χ^2 difference	*p*
Number of attempts			
Full model	1064.8	15.5	< .001*
GLM model without random effects	1078.3		
Time taken			
Full model	1320.5	21.7	< .001*
GLM model without random effects	1340.2		
Time per attempt			
Full model	235.8	76.6	< .001*
GLM model without random effects	310.4		
* *Below significance threshold of .05*			

Appendix E

Model fit comparisons for Experiment 5

Source	AIC	χ^2 difference	*p*
Number of attempts			
Full model	1432.1	23.3	< .001*
GLM model without random effects	1453.3		
Overall time taken			
Full model	1981.9	20.7	< .001*
GLM model without random effects	2000.5		
Match selection time			
Full model	208.3	66.4	<.001*
GLM model without random effects	272.7		
Mismatch time			
Full model	178.0	8.93	.0028*
GLM model without random effects	184.9		
Match delay			
Full model	243.4	49.7	<.001*
Model without random effects	291.1		
* *Below significance threshold of .05*			

Appendix F

Calculation of accuracy-based and time-based goals for Experiments 6 and 7

A pilot study was conducted with three participants to determine the targets used in the experimental conditions in Experiment 6. The mean age of the pilot sample was 44.6 years, and two of the pilot participants were male. The pilot participants completed ten rounds of the control motorcycle Pelmanism task. The procedure was identical to that of the *control* condition in Experiment 6.

The accuracy-based goals used in the experimental conditions in Experiment 6 were based on the number of attempts made by the pilot participants. The mean number of attempts and the maximum number of attempts were the benchmarks for the hard and easy accuracy goals in the first round respectively. To increase the level of difficulty for subsequent rounds, two regression lines were then calculated to determine how much the goals should decrease with each subsequent round. The regression lines resulted in a decreased in 1 attempt for every subsequent hard goal and a decrease in 4 attempts for every subsequent easy goal. The calculation is demonstrated below.

The time-based goals used in the experimental conditions in Experiment 7 were based on the time taken by the pilot participants. Similar to the accuracy-based goals, the mean time taken and the maximum time taken were the benchmarks for the hard and easy goals in the first round respectively. The increase in level of difficulty in goals for subsequent rounds were also calculated in the same way. The regression lines resulted in a 5 second decrement for every subsequent round for both hard and easy goals.

Taking Y = number of attempts or time taken, β = decrease in every subsequent round, x = Round; $1 \leq x \leq 10$:

For the hard goals:

$$Y_h = -\beta_h(x - 1) + intercept_h$$

$$\text{where } \beta_h = \frac{x1\ mean - x10\ min}{10} \text{ and } intercept_h = x1\ mean$$

For the easy goals:

$$Y_e = -\beta_e(x-1) + intercept_e$$

where $\beta_e = \frac{x1\ max - x10\ mean}{10}$ and $intercept_e = x1\ max$

Appendix G

Model fit comparisons for Experiment 6

Source	AIC	χ^2 difference	*p*
Motivation			
Full model	1821.3	361.0	< .001*
GLM model without random effects	2180.3		
Number of attempts			
Full model	4668.1	197.9	< .001*
GLM model without random effects	4863.9		
Time taken			
Full model	176.4	294.1	< .001*
GLM model without random effects	468.5		
* *Below significance threshold of .05*			

Appendix H

Model fit comparisons for Experiment 7

Source	AIC	χ^2 difference	p
Motivation			
Full model	2272.0	448.7	< .001*
GLM model without random effects	2718.7		
Time taken			
Full model	224.9	340.7	< .001*
GLM model without random effects	563.5		
Number of attempts			
Full model	6307.4	281.6	< .001*
GLM model without random effects	6587.0		
* *Below significance threshold of .05*			

Appendix I

Model fit comparisons for Experiment 8

Source	AIC	χ^2 difference	*p*
Search times – Visual search task			
Full model	3805.6	755.3	< .001*
GLM model without random effects	4556.9		
Detection accuracy – T-junction task			
Full model	5491.4	1500.3	< .001*
GLM model without random effects	6977.7		
Identification accuracy – T-junction task			
Full model	2014.0	191.4	< .001*
GLM model without random effects	2201.4		
* *Below significance threshold of .05*			