UNDERSTANDING HAZARD PERCEPTION IN FILMED AND SIMULATED ENVIRONMENTS

Patrick Runham, MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

2016

Abstract

Each year millions of people around the world are killed or are injured due to being involved in collisions while driving on the roads, with young and inexperienced drivers found to be significantly more likely to be killed or injured than older and more experienced drivers. Numerous studies have found a link between the likelihood of a driver being involved in a collision and their hazard perception (from this point on referred to as HP) ability, with young novice drivers having inferior HP abilities compared to older and more experienced drivers. As a result of these findings, governments in a number of countries have implemented or are investigating implementing HP testing elements into their driver examination programmes. This thesis presents a series of studies that investigate the factors that affect HP performance as well as comparing different HP testing methods. The more traditional video-based method is explored as well as new methods utilising a high-fidelity 180 degrees field-of-view driving simulator (used as an analogue of real world driving) in order to see if there are ways of making HP testing more representative of detecting and responding to hazards while driving on the road. This thesis also explores the use of functional near-infrared spectroscopy (fNIR) as a portable and flexible means of measuring dorsolateral prefrontal cortex (DLPFC) activity during driving.

In study 1 a multidimensional scaling analysis was carried out upon a set of data from the official UK HP test in order to explore if there were any underlying factors accounting for test performance in the data. No underlying factors were uncovered, indicating that the individual hazard videos used in the test are generally testing the same functions and that the sets of video hazards used in the test are well balanced. Study 2 moved onto comparing different HP testing methods by creating simulated driving hazards (in which participants were required to drive through simulated scenarios that each contained driving hazards) to compare with video hazards (that the participants watched on a T.V.) that are very similar to those used in the UK HP test. The results of the study revealed significant differences in behavioural psychophysiological responses between the simulated and video hazards, indicating that that the video-based HP testing method may not be representative of HP while driving on the roads and that

introducing simulated elements into HP testing may provide a more realistic test of HP abilities.

Study 3 consisted of an fNIR validation study carried out using three classic frontal tasks: the Wisconsin Card Sorting task, the Controlled Oral Word Association task and the Corsi Block task. The results of the study indicated that the fNIR system was primarily recording task workload and potentially task inhibition. Study 4 used fNIR to record DLPFC activity during simulated overtaking and following driving tasks, with the results revealing significantly higher DLPFC activity during the overtaking tasks. This DLPFC activity appeared to be more closely related to increased task workload rather than the inhibitory elements of the task. Study 5 consisted of an altered version of study 4, with the effect of driver age, driving experience and gender on driving behaviour and DLPFC activity being studied. The results again revealed significantly higher DLPFC activity for overtaking compared to following, however no effect of age, driving experience or gender on DLPFC activity was found. Males were found to overtake more times than females though. Study 6 brought the research regarding HP testing and fNIR together by recording DLPFC activity as participants took part in the same simulator based HP test used in study 2 or a replay HP test in which they simply watched and responded to replays of the simulated hazards. DLPFC activity was found to be similar in both conditions, indicating that the replay method may provide a relatively affordable and practical means of bridging the gap between traditional video-based HP testing methods and HP while driving on the roads. These findings have important implications for HP testing and the use of fNIR in driving research.

Table of Contents

List of Figures

List of Tables

Chapter 1- General Introduction

1.1 Thesis Outline

The research in this thesis was designed to address a selection of both applied and theoretical questions about HP while driving. These include both practical questions about the use of HP testing as part of the current UK driving test and theoretical questions about more general aspects of driver behaviour in dangerous situations as experienced through videos or simulations of hazardous driving situations.

So far there has not yet been a systematic published study of data from the UK HP test, including what elements of the test produce the most benefits. Therefore one of the major aims of this thesis is to carry out an analysis of data from the UK HP test in order to understand what aspects of it are best at testing HP abilities. In particular, the analysis will look for the hazards (and the properties of those hazards) that are best at discriminating between test candidates who fail the HP test and those that pass. This will involve the creation of a HP typology. This information could help inform the Driver and Vehicle Standards Agency (DVSA) if there are any areas in which the test could be improved, which in turn may help further improve road safety in the UK. Having a better theoretical understanding of the HP test and of drivers' HP skills more generally will help the road safety community internationally in interpreting driver behaviour in hazardous situations, testing it and improving it

An important issue with the UK HP test and video-based HP tests in general is whether or not a video-based test is representative of real world HP. This thesis will also investigate this issue by directly comparing video hazards from the UK HP test with simulated versions that will be created in a dedicated driving simulator. This will also help explore the feasibility of introducing any simulated elements into the test for more realism. A further issue that will be explored relates to the psychophysiology of driving. Psychophysiological recordings during HP and other driving tasks will be taken to investigate this issue. The role of the frontal lobes, specifically the dorsolateral prefrontal cortex (DLPFC), in driving will also be explored using functional near-infrared spectroscopy (fNIR), a relatively new neuroimaging technology that has exiting applications for driving research. This will involve carrying out a pilot study with the fNIR device to better understand what functions it is measuring and then using it to explore the role of the DLPFC in HP and general driving. The effect that individual differences such as driving age and experience have on driving performance will also be investigated. Combined, this research will allow for a better understanding of what influences HP and driving abilities, which in turn can be used to improve road safety.

1.2 Overview

Every day, millions of people around the world drive vehicles on public roads. In the modern world driving has become an integral and indispensable part of people's lives and the global economy. As a result of this the number of cars in the world surpassed 1 billion in 2011 and that number is increasing each year. Unfortunately, traffic crashes are one of the most common causes of death throughout the world. A report by the World Health Organisation revealed that more than 1 million people were killed on the roads in 2013. While the number of deaths on the road tends to be much higher in less developed countries, traffic deaths are still a major problem in more developed countries. There is a particular problem with young and inexperienced drivers being killed or seriously injured on the roads. The personal and economic costs of these road deaths are enormous. It is therefore imperative to explore ways to improve road safety in order to reduce the number of deaths that take place on the roads. One area of driving that has been researched heavily in recent years with regards to road safety is HP. This research has led to countries such as the United Kingdom introducing hazard perception tests into their national driving tests, with the aim of improving road safety by improving the hazard perception abilities of newly qualified drivers.

The purpose of this chapter is to provide a review of the hazard perception literature. This will involve introducing the concepts and theories of driving hazards and hazard perception, as well as covering the development of hazard perception testing and the findings revealed through hazard perception testing. The development, current form and impact of the UK hazard perception test will also be discussed. This chapter will also provide an overview of the research aims of this thesis.

1.3 Hazard Perception

1.3.1 Driving Hazards

Before exploring what HP is, it is important to first define what a driving hazard is. There is no agreed upon definition for a driving hazard and as a result a number of different definitions have been used in the driving literature. Benda and Hoyos (1988) defined a hazard as “the possibility that a mass, i.e. a vehicle, might undergo a change in velocity or direction by colliding with a moving or non-moving object or by swerving off the road” (p.1). Wallace, Haworth and Regan (2005) defined hazards as “any permanent or transitory, stationary or moving object in the road environment that has the potential to increase the risk of a crash” (p.12). Mills, Rolls, Hall and McDonald (1995) define a hazard as “any aspect of the road environment or combination of circumstances which exposes an individual to an increased possibility of an accident” (p.1). Meanwhile Graham and Kinney (1980) referred to hazards as “some potential danger beyond one’s immediate control” (p.13). McKenna and Crick (1994) referred to a hazard as a situation “in which you might consider it necessary to take some form of evasive action, by braking or steering etc.” (p.28). Lerner and Rabinovich (1997) took a slightly different approach and introduced the concept of perception into their definition of a hazard. They defined a hazard as situations that pose an increased risk to a driver even if they do not perceive the risk, as opposed to situations in which the driver may perceive a risk even though it does not actually exist.

Despite the large number of definitions for driving hazards, there are a number of common themes and similarities that run through them. Haworth, Symmons and Kowaldo (2000) carried out a review on the definitions of driving hazards that had been used in driving research. Based upon previous definitions they defined a hazard as “any combination of road environment situation and/or circumstances that increases the probability of a crash, and is beyond the immediate control of the driver in question” (p.10). Based upon this definition they stated that hazards include: “semi-permanent physical characteristics of the road environment, temporary physical characteristics of the road environment and behavioural aspects of the surrounding traffic” (p.10). This definition encompasses the condition of the road surface, objects in the road, other vehicles in the road as well as the weather conditions. These objects and conditions may be transitory or permanent but one thing they have in common is the potential to cause a crash.

1.3.2 Hazard Perception

As is the case with hazards, there is no universal definition of hazard perception and a number of different definitions have been used in the driving literature. McKenna and Crick (1994) stated that HP is "the ability to recognise dangerous road and traffic situations, or the ability to read the road" (p.28). In a similar manner to this, Mills et al. (1998) defined HP as "the ability to identify potentially dangerous traffic situations as they arise or the ability to 'read the road' and is a skill which is directly to the driving situation" (p.2). McKenna (2000) stated that, "By hazard perception I mean the ability to anticipate, the ability to read the road" (p.1). Deery and Love (1996) defined hazard perception as "the process of identifying hazardous objects and events in the traffic system and quantifying their dangerous potential" (p.2). Sagberg and Bjornskau (2006) defined HP in relation to two processes, "the ability to anticipate traffic situations, of which there are two separable components; one is the degree of the perceived hazard associated with the situation and the other is the perception-reaction time to the perceived hazard" (p.1). Crundall, Chapman, Trawley, Collins and van Loon (2012) defined HP as "the process of detecting, evaluating and responding to dangerous events on the road that have a high likelihood of leading to a collision" (p.1). This definition introduces the idea HP might be divided up into three sequential stages - detecting a hazard, evaluating it and then responding to it.

Although there are some common themes running through many of these definitions of HP, there are also some significant differences. For example, some of the definitions involve the process of decision-making and taking actions, where others focus entirely on the process of perception and detection. The failure to have a standard definition of HP is problematic and leads to issues in HP research. What is clear from these definitions is that HP concerns detecting potentially hazardous situations in the environment and determining strategies and responses to avoid those hazards.

A number of theories and frameworks of HP have been proposed. HP has often been viewed in the context of Situation Awareness (SA). The most widely accepted description of SA comes from Endsley (1995) who defined SA as “the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p.35). Endsley (1995) proposed a hierarchical three level model of SA, with one level leading to the next. The model comprises: Level 1, the perception of elements in the environment; Level 2, the comprehension of the current situation; Level 3, the projection of future status. Achieving all of these elements in necessary to achieve SA. Level 1 refers to “perceiving the status, attributes, and dynamics of relevant elements in the environment” (Endsley 1995, p.35). Level 2 refers to going “beyond simply being aware of the elements that are present to include an understanding the significance of those elements in light of pertinent operator goals” (Endsley 1995, p.36). Level 3 refers to “The ability to project the future actions of the elements in the environment” (Endsley 1995, p.36).

When looking at the definitions of HP, it is clear that the elements of perception, understanding and projection in Endsley’s model of SA are applicable to HP. Level 1 could refer to a driver perceiving other vehicles as well as properties such as the speed they are travelling at, as well as environmental factors such as the road conditions. Level 2 could refer to a driver combining information regarding other vehicles on the road and the road conditions in order to comprehend how these factors may affect you as a driver. Level 3 could refer to a driver hypothesising and anticipating the potential actions of other vehicles on the road in those conditions and their consequences.

From this process the driver must then make a decision as to what action they must take. However Endsley (1995) emphasised that the decision-making process, although heavily influenced by SA (i.e. good decision making is dependent upon using SA effectively to make fast and valid predictions), is a separate process. Endsley (1995) argued that SA is like other skills, in that with practice, it becomes an automatic process. With regards to HP this means that through the process of encountering hazards on the road, over time a driver will automatically be able to utilise the different aspects of SA. Endsley (1995) stated that sufficient practice would result in result in the formation of mental schemata in long-term memory. When these schemata have formed, a driver could automatically compare the elements of a driving situation to those in memory, instantly accessing relevant schemata and their associated actions. This is opposed to going through each level of SA and then deciding upon which action take. This extra time may be the difference between making a correct decision and a wrong decision, as well as making a decision in time. This suggests that experienced drivers will be able to use SA to make predictions regarding traffic situations quicker and with more ease than inexperienced drivers. This in turn will allow them to make decisions more quickly than inexperienced drivers, potentially allowing them to avoid crashes inexperienced drivers cannot.

SA also interacts with workload to affect the ability to process information, which in turn affects performance (Endsley, 1995). High levels of SA and low levels of workload will result in information being processed easily, whereas high SA combined with high workload will result a person having to work hard but ultimately being able to process necessary information, allowing a task to be performed well. Conversely, low levels of SA and high levels of workload will lead to there being too much information to process, whereas low levels of SA and workload will result inattentiveness and poor vigilance, which in turn will lead to poor task performance. With regard to HP, low levels of SA or high levels of workload could lead to a person missing a hazard and crashing as a result.

Fitzgerald and Harrison (1999) proposed that HP is a multifaceted skill dependent on three cognitive and behavioural components: workload management, automation and attention. Workload management refers to a driver having enough cognitive capacity to for vehicle control and navigation as well as non-driving activities. While experienced drivers have cognitive strategies for dealing with high levels of workload, inexperienced drivers do not which may lead to cognitive overload and limited HP abilities. The authors argued that while the use of a car's controls becomes an automatic process, hazards are encountered infrequently meaning that novice drivers may not yet have developed appropriate responses to them. With regards to attention, the authors stated that this is critical, as drivers need to be able to switch attention dynamically (to scan the driving environment for example). They argue that without these abilities, HP will not function properly.

Fitzgerald and Harrison (1999) explained HP in the context of the recognition primed decision-making (RPD) model of Klein et al. (1989, 1993). RPD involves several steps based around deploying attention to a situation and ultimately producing a correct response to that situation. The first step in the process is 'situation recognition', in which the current situation is compared to those in memory to establish if the situation is novel or familiar (i.e. has this situation been encountered before). If the event is classified as familiar then previous responses and outcomes can be evaluated to determine if they will help in the current situation. From this a list of potential responses is generated.

The next stage is the 'serial option evaluation', which involves taking the list of potential responses and testing them in order to work out their consequences and which is the best response to make. The important part of this stage is prior experience. If a person has experienced a similar situation before then that can increase their chances of having an optimal response, and if the previous response was unsuccessful then it can be ruled out. If a person has not experienced a similar situation before then they will not have these advantages. For example, if a person has previously dealt with a car suddenly pulling out from a side road they may already know a strategy to deal with that (the particular braking and steering inputs required etc.). Another important factor is the similarity of the previous experience. While having encountered an identical situation will allow for the production of an optimal response quickly, having only encountered similar situations will result in the need to test responses based on those previous situations. For example, if a car pulls out suddenly and this time the weather conditions are icy as opposed to clear, this will impact on the decision making process as the previous braking and steering controls may not work in this current situation. This in turn may lead to slower reaction times and an increased likelihood of crashing.

Unlike other models of HP, this places more of an emphasis on the decision-making and response to hazards. Fitzgerald and Harrison (1999) stated that the general view of HP only encapsulates the situation recognition process of RPD. They argued that there should be a focus on 'hazard behaviour', which encompasses the perception of hazards as well as the process of making appropriate responses in order to avoid crashing. According to Fitzgerald and Harrison (1999), perceiving a hazard does not in itself allow a driver to avoid having a crash; a suitable response must be achieved as well. The authors argues that by viewing HP in terms of both perception and response, it allows for a better understanding of what must be achieved to avoid crashes, and RPD is an appropriate model for this.

1.3.3 Video-Based Hazard Perception Testing

The testing and research of HP spans several decades and has involved a wide range of different techniques and methodologies. Pelz and Krupat (1974) pioneered HP testing by having participants watch and respond to videos that contained driving hazards. They had participants watch a 5-minute film (taken from the perspective of a car driver) of driving within a city. The film contained a total of 10 hazardous events. The film was projected onto a screen that the participants sat in front of. The participants responded via a continuous measure by moving a lever across a meter that ranged from 'safe' to 'unsafe' (with 10 gradients in-between). The lever response indicated how safe or unsafe they felt while watching the driving scenes. The participants were told that if they felt at ease while watching the film (when the car in the film was parked or at rest for example) then they keep the lever at the 'safe' end. However, if they felt threatened, in danger or at risk of having a serious accident then they should move the meter all the way to the 'unsafe' endpoint. The participant was viewed to have responded to a hazard when the caution level (how safe or unsafe they felt) increased (towards unsafe). This was called the 'point of onset". The point at which the caution level fell to its original position was referred to as the "point of offset".

The results revealed an association between the crash liability of the participants and the levels of apprehension they displayed via the continuous lever response. Drivers who had fewer accidents were found to be more cautious (i.e. they tended to put the lever in an 'unsafe' position more of the time indicating that they tended to perceive greater risk) and to respond more quickly to the onset of the hazards. Interestingly, four of the hazards in the film were not picked up by many of the participants, as there were very few recorded onset or offset points. The authors stated that these hazards were potentially ambiguous.

A further development in HP testing and research came from Watts and Quimby (1979) who created a simulated driving environment. They used the same method of presenting participants with film of on-road driving taken from the driver's perspective. However, they made significant changes to the testing environment by having the participants sit in the driver's seat of a shortened car body. The participants were also provided with a rear view through the inclusion of views of the rear-view mirror and right-hand wing mirror. Road and engine noises recorded when filming the driving scenes were included to provide realistic levels of noise and vibration. Indicator lights also switched on and off when the car in the film changed lanes. The purpose of these changes was to provide extra realism while testing. A potential criticism of the method used by Pelz and Krupat (1974) is that the HP testing environment they used was not representative of HP while driving. In order to validate their rudimentary driving simulator, participants drove a 16-mile route and then watched a film of a drive on that 16-mile route in the simulator. During both driving tasks, the participants were required to provide scores relating to the level of risk they perceived at 45 preselected locations. An 11-point scale was used, with a score of 0 indicating that there was no chance of having an accident or near miss, whereas a score of 10 indicated that there was a high chance of having an accident. The results revealed that scores in the simulator correlated with those from on-road driving.

Carrying on from this work, Quimby and Watts (1981) used their driving simulator to test HP. The same method that was used to validate the simulator in the previous study was utilised. However this time the participants were also required to list any hazards that they had encountered during the on-road segment. Meanwhile their reaction times to seven hazards in the filmed segment were recorded using the participant's continuous responses. Driving performance was also measured and the participant's accident histories were recorded. They found that the participants overall level of perceived risk was associated with their accident liability. Participants with a higher overall level of perceived risk had a lower accident liability. This is in line with the findings of Pelz and Krupat (1974). The results also revealed that the participants who reported more hazards made fewer driving errors. No relationship between HP response times and accident liability was found for the filmed hazards however. The authors attributed this to some of the hazards not being appropriate as a large number of the participants simply failed to notice them. When these hazards were removed from the analysis, a small but significant correlation between reaction times and accident liabilities was revealed, with slower reaction times associated with higher accident liabilities. This is similar to the issue encountered by Pelz and Krupat

(1974) and indicates how strictly the content of HP videos must be controlled in order to adequately test HP performance. The study also revealed an effect of age, with younger drivers displaying slower HP reaction times than older drivers. The authors attributed this to inexperience on behalf of the young drivers as their reaction times during a reaction time task were actually faster than the older drivers.

A major change in HP testing and research took place as a result of the work of McKenna and Crick (1994). Importantly, they did not believe that is was necessary to have participants sat in a car as the most important part of a HP test are the driving video clips themselves. They thought that having car controls in a simulator (that could not actually be used) contributed to 'redundant realism'. They also replaced a continuous lever response with a discrete button response, to be pressed when the participant perceived a hazard. McKenna and Crick (1994) defined a hazardous situation (i.e. a driving hazard) as "one in which you might consider it necessary to take some form of evasive action, by braking or steering etc." (p.28). Projector film of on-road driving with hazardous events was replaced by videos (also taken from the driver's perspective) that contained predetermined hazard onset signals - times that marked the point at which a hazard became visible. This signal switched on a computer timer, which stopped if the participant noticed the hazard and responded with a button press. If there had been no response after a certain period of time, the computer stopped timing and the participant was deemed to have missed the hazard. A further difference is that McKenna and Crick did not compare drivers who differed in accident liability, rather drivers who differed in their level of driving experience. This is an extremely important difference as the validation of the testing method and the hazard video clips was based upon identifying drivers who differ in levels of experience as opposed to crash liability. This emphasizes the role of HP testing as a way of identifying novice drivers rather than risky ones and stresses the idea that HP may be a skill that is acquired with practice rather than a personality difference. The results of the study revealed that experienced drivers responded faster than novice drivers when detecting hazards.

Mckenna and Crick (1994) further refined this method of HP testing by validating new video hazards as well as altering the timing of the scoring windows. The results again revealed that inexperienced drivers had slower reaction times to hazards than experienced drivers. The authors noted that certain hazards were better at discriminating between experienced and inexperienced. Hazards that required the participants to look far ahead were particularly, while hazards in which the starting point was ambiguous were poor at discriminating driving experience. These findings have been reproduced by other, more recent, studies (McGowan and Banbury, 2004; Wallis and Horswill, 2007). The study also found that HP reaction times differed on the basis of expertise and training. Experienced drivers were found to have slower HP reaction times than police officers that had undergone advanced driver-training courses. This strongly suggested that HP abilities might be improved by specific training.

These findings indicate that hazard perception can discriminate on the basis of driving experience. However, other studies have not reproduced these findings. Chapman and Underwood (1998) used a very similar design to studies that had found differences in response times between experienced and inexperienced drivers. Young novice drivers (who had only passed their driving tests within the previous 3 months) and older experienced drivers (who had possessed a driving licence for between 5 and 10 years) took part in the study. They watched 13 video clips (there were 39 total video clips that were divided into 3 comparable groups) that contained between 1 and 4 driving hazards and had to respond with a button press when they noticed each hazard. Timing windows, corresponding to 1 second before the hazardous event began until the objects involved disappeared, were used to calculate the participant's HP reaction times. These windows ranged from 2 seconds to 16 seconds. The authors also carried out a pilot study to make sure that the hazards in the videos were detectable.

The results failed to reveal differences in HP response times between experienced and inexperienced drivers. There were also no significant differences in the number of button presses or the number or hazards the participants reported having identified. The authors argued that the results might have been due to the hazards in the video clips being sufficiently well defined that both experienced and inexperienced drivers easily detected them. They also argued that the slower reaction times of the older participants among the experienced group (the more experienced drivers tended to be older than the inexperienced drivers) might have counteracted the advantages in reaction times as a result of greater experience amongst the group. Crundall, Underwood and Chapman (2002) used a similar design to Chapman and Underwood (1998) and also failed to find differences in HP reaction times or the number of hazards responded to as a function of driving experience. They made a slight change to the methodology by having participants respond to hazards via a foot-pedal that produced millisecond accurate responses.

Sagberg and Bjornskau (2006) tested HP using video-based hazards. They tested the HP reaction times of inexperienced drivers who had held a licence for 1, 5 or 9 months as well as experienced drivers who had held a licence for at least several years. The drivers were shown 20 minutes of video of naturally occurring driving scenes (taken from the perspective of the driver). The participants had to respond to 'critical events' that were defined as "as any motion by some other road user, which could possibly develop into a hazard, and for which the driver had to be especially prepared for taking some evasive action in terms of braking or steering". In total there were 31 critical moments that were scored on the basis of responding within a predetermined response window. These windows ranged from 4s to 25s. The beginning of the response window was defined by the point in time when a potentially hazardous motion by another road user could be detected. The end of the response window was defined by the last possible time in which a braking or steering response would have an effect. The results revealed no significant differences in the HP response times between experienced drivers and novice drivers. This is despite the fact that the experienced drivers had been driving for an average of 27 years while the novices had, at most, 9 months experience.

There are potential explanations for these contradictory HP findings. Many of the studies listed above used HP videos that contained different driving scenes and hazards. This fits with the findings from Pelz and Krupat (1974) as well as Watts and Quimby (1979), who found that only certain filmed driving hazards were capable of discriminating accident liability. This is further supported by the fact that Chapman and Underwood (1998) argued that the hazards in their videos might have been too easy to detect, and therefore unable to differentiate experience. Meanwhile McKenna and Crick (1994) reported that hazards that required the participants to look far ahead were the best at discriminating levels of experience, whereas hazards that had ambiguous onsets were poor at discriminating levels of experience. Sagberg and Bjornskau (2006) argued that anticipation, surprise and complexity were the most important dimensions of successful HP clips they developed.

Underwood, Crundall and Chapman (2011) argued that some hazards are abrupt and attention gathering (e.g. a pedestrian suddenly stepping into the road). As a result these hazards are potentially unavoidable and may not be able to discriminate experience levels as the hazards capture attention regardless of experience. Meanwhile hazards in which the onset is gradual and require drivers to anticipate what might happen (utilising level three SA) if another road user behaves in a hypothetical way (e.g. realising that if a car pulls out from a side road ahead you will have to slow down) are better able to discriminate levels of experience. Crundall et al. (2012) also proposed that part of the problem might be that hazards employed by different studies are not comparable in their properties.

1.3.4 Simulator-Based HP Testing

A more recent development in HP testing involves the use of driving simulators. While Watts and Quimby (1979) attempted to simulate being inside a car while testing HP, this was not a driving simulator in the true sense as the car controls were not operable. With the rapid advance of computer technology, it has become possible to create driving simulators that provide high levels of realism. The use of driving simulators provides a number of advantages. For example, a driving simulator allows researchers to create strictly controlled driving scenarios and hazards. This is not always possible when filming on public roads, as it is not possible to control the traffic on the roads. Given the research that has indicated how important the properties of hazards are for HP testing, this is a huge advantage. Simulators can also provide a much larger field of view (similar to that of actually being in a car) than showing a video on a computer screen as well as providing auditory and tactile feedback. But perhaps the most important advantage is that driving simulators provide the ability to control a simulated car with actual car controls (e.g. steering wheel, brake pedal) and interact with the driving environment. This allows participants to respond to hazards as they would on the road, by braking or changing direction for example. This provides a more realistic means of testing HP than simply pressing a button, allowing experimenters to collect information on the driving behaviours of participants, as well as other measures such as eye movements.

Liu, Hosking and Lenne (2009) used a motorcycle simulator to investigate differences in HP between experienced and novice riders. The motorcycle simulator consisted of: handlebars, a saddle, a gear lever and foot brake controls. It was designed to simulate a 400cc motorcycle. The simulator allows riders to simulate using front and rear brakes, throttle and steering. The participants were divided into 4 groups on the basis of their riding and driving licences: experienced motorcycle riders with a full driver licence; inexperienced motorcycle riders (learner licence) with a full driver licence; novice motorcycle riders (no learner licence) with a full driver licence; novice motorcycle rider (no learner licence) with a learner driver licence. The participants rode through 3 motorcycling scenarios: riding in a rural area, riding in a Central Business District with moderate traffic, riding in a central business district (CBD) with high levels of traffic. Each of the scenarios contained 8 hazards, including hazards with pedestrians, treacherous road surfaces, other vehicles on the road and difficult bends in the road. The simulator allowed the experimenters to collect measures of speed, steering, acceleration, braking, lane position and the number of crashes. The participants were evaluated in terms of the number of crashes related to hazards, their speeds on approach to hazards and an overall driver rating.

The lowest skill group (novice riders and inexperienced drivers) were found to crash more in the rural scenario. However there were no differences among the other 3 groups. As a result the authors believed that the number of crashes in the simulator might have not been a useful measure of experience or riding ability for this scenario. The authors also noted that that the number of crashes was very high in general. This may have been due to the participants' unfamiliarity with riding in the simulator or a sign that the simulator was simply unrealistic. For the medium traffic density CBD scenarios, experienced riders were found to score significantly higher than the novice drivers. For the medium traffic density CBD scenario, experienced riders were found to approach hazards at lower speeds than novice riders.

There was also some evidence that approach speed to hazards were related to riding experience for some hazardous road conditions. Interestingly, the authors noted that sudden onset hazards that appeared suddenly and did not involve any anticipation were not good at discriminating experience. Shahar, Poulter, Clarke and Crundall (2010) used the same motorcycle simulator to see if they could discriminate between motorcycle riders and car drivers (who did not have experience of riding motorcycles) in terms of their responses to hazards. While it was possible to differentiate between the groups on the basis of their driving during safe periods (in terms of their handling etc.), it was not possible to do this in terms of their responses to hazardous situations. Although they used the same scenarios as Liu et al. (2009), they did not use the same hazards. This may explain the difference in findings between the two studies.

Shahar et al. (2010) proposed that the failure of the hazards to discriminate between riders and drivers was due to a lack of precursors or foreshadowing to the hazard. The idea of hazard foreshadowing is described in Garay-Vega, Fisher and Pollatsek (2007). Foreshadowing is essentially a cue or precursor to a hazard. Foreshadowing may be a non-hazardous event that draws a driver's attention to the subsequently occurring hazard. For example, a person getting into a parked car in the road ahead provides a cue that the car may pull out so special attention needs to be paid to the car. The idea is that experienced drivers will be able to use the foreshadowing of hazards to anticipate that a hazard may occur. The discriminatory effects of these hazard precursors may relate to experienced drivers better visual scanning of driving scenes (Pradhan et al., 2005; Garay, Fisher and Hancock, 2004). Shahar et al. (2010) stated that several of their hazards lacked the requisite amount of foreshadowing to allow the experienced riders to demonstrate any superior HP skills.

These findings indicate that the properties of the hazards used in HP tests are vitally important, as the inconsistencies in video-based HP testing have also been found in the simulated environment. In order to discriminate factors such as experience, the hazards must have certain properties or they will not be ale to discriminate these properties. One of the major factors appears to be associated with the levels of anticipation involved in a hazard, which may in turn be liked to level 3 of Endsley's (1995) model of SA. Another issue raised is that the design of the simulators is of paramount importance. If the simulators are not providing a realistic driving experience then they may not be able to properly test driving abilities such as HP. This makes the design and validation of simulators and hazards extremely important. Fortunately, simulators provide researchers with the flexibility to investigate the properties of hazards.

Crundall et al. (2012) addressed the issue of hazard design by utilising the flexibility of driving simulation to create simulated hazards that were defined by the relationship between the hazard and the precursor to the hazard. In order to design successful hazards, they reviewed the literature on what constitutes a successful HP clip. They noted the findings of Sagberg and Bjornsaku (2006), who reported that anticipation, surprise and complexity were important, as well as those of Chapman, Underwood and Roberts (2002) reported that anticipation, knowledge and scanning are vital processes undertaken during HP.

Anticipation can be equated to Level 3 of the SA model of Endsley (2000). In this context anticipation is viewed as the use of one's perception (Level 1) and comprehension (Level 2) of the environment to project the future outcome of that environment (Level 3). Knowledge (or experience) refers to the having the ability to associate certain stimuli as being likely to occur together. In relation to the model of Sagberg and Bjornskau (2006) it refers to a driver being surprised by an event having failed to link the precursor and the hazard.

Sagberg and Bjornskau (2006) used the idea of complexity to describe the need for a driver to analyse multiple sources of potential hazard in a complex driving environment. It has been shown that drivers must employ wide scanning strategies in order to drive in a safe manner (Mills, 2005). For example, novice drivers have limited scan patters and tend to focus on small areas of the visual scene (Crundall and Underwood, 1998), which may result in their relatively poor HP abilities and increased propensity to be involved in accidents. Crundall (2009) reported that novice drivers have difficulty dividing and focusing attention. Crundall et al. (2012) argued that although dividing attention across a number of potential hazards is vitally important, when a hazard actually occurs it is focused attention on the hazard that is required.

From these ideas the authors designed three types of hazard. Behavioural prediction (BP) hazards, environmental prediction (EP) hazards, as well as dividing and focusing attention (DF) hazards. BP hazards are those in which the precursor and the hazard are the same stimulus. For example, a car with its engine running is parked at the side of the road (the precursor) when it suddenly pulls out in front of you (the hazard). In this instance there is a direct link between the precursor and the hazard, allowing the prediction of a future event (the hazard) directly from current behaviour (the precursor) through anticipation.

EP hazards are hazards in which the precursor and hazard are different stimuli (knowledge). An example of an EP hazard is an occluded pedestrian (the hazard) stepping out from in front of a bus parked at a bus stop (the precursor). In this case the precursor and hazard are indirectly linked, requiring the driver to use their experience and knowledge to predict that two distinct stimuli (the precursor and the hazard) can occur together.

DF hazards reflect a driver's scanning of the driving scene, dividing attention across multiple targets and then focusing on a single target when a hazard has occurred. A DF hazard contains more than one precursor (BP or EP). The drivers must divide their attention between the two precursors and then focus their attention on the precursor that develops into a hazard. An example of a DF hazard would be a bus parked at the side of the road obstructing the view ahead (EP precursor) meanwhile a pedestrian is walking on the opposite side of the road (BP precursor). The pedestrian then crosses the road forcing the driver to focus attention on the pedestrian crossing the road.

Simulated HP clips were then created that contained BP, EP or DF hazards. Drivers of varying experience (learners, experienced drivers and driving instructors) took part in a driving simulation study in which they had to respond to the BP, EP and DF hazards. Their eye movements were monitored while they drove in a medium fidelity fixed-base driving simulator.

The results of the study revealed a variety of effects of experience level on driving performance. Learner drivers made no distinction between precursors and hazards with their fixations and were equally poor at fixating both precursors and hazards compared to experienced drivers and instructors. The fact that all of the drivers fixated less on the less salient precursors indicates that drivers do not utilise all of the available cues to help them detect hazards. Learner drivers were more likely to miss BP precursors than the other drivers but just as likely to spot the BP hazards. The opposite effect was found in the EP conditions, with leaners just as likely to spot the precursors but more likely to miss the hazards.

The authors explained that the direct link between BP precursors and hazards would increase the likelihood of drivers fixating on BP precursors compared to EP precursors. This explanation was supported by the fact that experienced drivers and instructors fixated on BP precursors more than EP precursors. Careful analysis of the data showed that although learner drivers were just as likely to spot the BP hazards, they are slower at doing so than experienced drivers and instructors. Experienced drivers and instructors increased likelihood of fixating the BP precursors may have led to their reduced times taken to fixate the BP hazards This result shows the importance of spotting BP precursors as it allows for quicker targeting of the actual hazards.

EP precursors were fixated less frequently than the other forms of precursors, with all drivers fixating them with equal probability. Despite this, learner drivers failed to detect as many of the EP hazards as the experienced drivers and instructors, and took more time to fixate than the other drivers. Without an EP precursor, the EP hazard is (as there is no direct link between the two) an abrupt onset. The authors posited that it might be expected for younger and inexperienced drivers to respond quickest and most frequently to the hazard. As this was not the case it was suggested that experienced drivers and instructors may have been accessing predictive information about the EP hazard that was unavailable to the learner drivers.

One potential answer is that experienced drivers are able to process more information within each fixation compared to inexperienced drivers, ultimately providing them with more information about a potential hazard. This is supported by the findings of Chapman and Underwood (1998) who reported that novice drivers fixated hazards for longer than experienced drivers due to them not being able to process the visual scene as quickly and efficiently as experienced drivers.

An alternate explanation proposed by the authors is that the experienced drivers and instructors monitored the EP precursors peripherally until they needed to fully process them. It is possible that as EP precursors have an indirect link with EP hazards (but increased accuracy of their predictive information), drivers need only be concerned with monitoring them if the hazard has a large probability of occurring. For example, a driver should only be concerned about pedestrians emerging from in front of a parked bus when driving past that bus. From this perspective, reduced fixations of EP precursors may be the result of drivers observing these precursors through their peripheral vision. A drawback to this strategy is that drivers may leave it too late to then fixate on the precursor, thus failing to react in time to the hazard itself.

1.3.5 On-Road Hazard Perception Testing

Realistic on-road HP testing and research has been limited due to the ethical and safety issues associated with testing driver performance in hazardous driving situations. There has however been some limited testing of HP on specially designed test tracks. This research has been carried out at Virginia Tech Transport Institute. They have built a 3.5km test track, consisting of a two-lane highway constructed to meet interstate standards. Their HP tasks have involved creating staged hazards to test driver's HP as they drive around the test track. They make use of an instrumented car.

Lee et al. (2008) tested newly qualified teenage drivers and adults on a test track that contained 3 hazardous driving scenarios: hidden stop sign (traffic lights are switched off and temporary stop sign is obscured by a van parked on the pavement), hidden pedestrian (maintenance worker steps in front of van and becomes obscured from view) and hidden pedestrian with lane closure (cones close off one lane and maintenance worker then proceeds to step in front of van parked in the closed lane and is obscured from view). The last of the scenarios also involved a distraction task in the form of a text-messaging task. The adult drivers were found to observe and recognise hazards more frequently than the young drivers. A large number of the young drivers were found to struggle with disengaging from peripheral tasks (i.e. the text messaging task).

In a follow up study, Pradhan, Simons-Morton, Lee and Klauer (2010) tested newly licensed teenage drivers in a combined HP and distraction task. These were the same participants who took part in the Lee et al. (2008) study, the aim being to explore the effect of experience on HP after 12 months of qualified driving. Participants drove around the same 3.5km test track as before, a total of 7 times. During the drive a number of driving hazards appeared. At the same time the participants were engaged in secondary distraction tasks. The hazards were the same as those used in the study by Lee et al. (2008) with the addition of a hazard which was similar to the hidden pedestrian and lane closure task, but this time the pedestrian was replaced with a mock dog, that was placed in front of the car, obscured from view. The distractions tasks were: an odometer task (reporting when the last digit was a 3, 6 or a 9), a texting task (writing out and sending a message supplied by the experimenter) and a mobile-phone task (dial traffic information services and verbally report on any incidents on major roads). For the texting and odometer tasks, the novice drivers displayed an improvement in HP. However, for the mobile phone task, none of the novice drivers displayed any HP or suspended the distraction task. The authors argued that this may have been due to the participants believing that as they were still viewing the road while they were on the mobile phone, they did not need to suspend the task as they thought they could detect hazards anyway.

1.3.6 Hazard Perception Training

One issue that has been researched with regard to HP is the ability to train and improve HP performance. Other driving skills such as vehicle control are trainable and provide the ability to give people driving lessons in order to prepare them to drive safely on the roads. If HP is a trainable skill, it will be possible to train and improve driver's HP skills. This in turn may improve the driving of newly qualified drivers and the safety on public roads.

McKenna and Crick (1994) observed that police officers who had gone through advanced driving training courses had lower response latencies compared to experienced drivers who had not gone through such training programmes. As a result of this they decided to directly test if driver training could improve HP performance. Drivers who had taken The Royal Society for the Prevention of Accidents (ROSPA) advanced driving course were compared with a control group who had not taken the course. The participants were matched to have similar ages and levels of driving experience. The ROSPA driving course consists of a series of lectures as well as on-road training with a driving instructor. The ROSPA course is not specifically designed to improve HP skills; rather it is aimed at improving a wide range of driving skills. HP was tested using a video-based test, in which the participants had to respond to hazards shown in driving scenarios through a response button. The participants underwent testing twice, once before and once after training. The results revealed that the drivers who went through the ROSPA driving course showed a reduction in HP response latencies

Mills et al. (1998) investigated whether or not it was possible to improve HP performance and what methods could achieve this. They tested 144 drivers who were between the ages of 17 and 25. Each of the participants was selected on the basis that they had recently passed their driving test (maximum of 3 months previously). The participants were then divided into four groups that underwent different HP training methods but the same testing methods. One of the groups was a control group (Group 1). The second groups received a classroom based HP training course that consisted of watching a video in the presence of a driving instructor, who provided them with information on scanning the scene and anticipating any developing hazards (Group 2). The third group received the same training as the second group as well as receiving a two-hour on-road training course from a driving instructor, who gave them training on how to identify hazards while driving (Group 3). They were also given feedback on areas that they needed to improve on. The fourth group received the on-road training course only (Group 4).

Before the training took place the participants completed a questionnaire that assessed driving attitudes and behaviours. They also completed a 14-mile long HP driving route during which they had to verbally identify any hazardous situations that arose. They were accompanied by a driving instructor, who observed them and recorded how quickly they identified the hazards and then responded to them. The instructor noted any errors and the location of these errors along the driving route, as well as providing an overall assessment in terms of safety, anticipation, observation and HP skill. The participants were also assessed via a video-based HP test that was based upon previous designs such as those of McKenna and Crick (1994). The video was taken from the perspective of a driver's view of the road. It lasted 15 minutes and contained 23 hazards. The participants had to respond to via a response button when they identified a hazard. These assessments were carried out before and after the participants had gone through their training programmes.

The results of the study revealed that the three training groups improved their HP scores on the video-based HP test, with the best improvements from group 3, followed by group 4 and then group 2. Group 3 was also found to have improved on the 14-mile HP driving route by increasing the proportion of hazards that they identified early and responded to correctly. The observing driving instructors noted that groups 3 and 4 performed best on their second attempt at the 14-mile driving route. There was also a positive correlation between performance on the video-based and driving-based HP tests (which the authors stated was a validation of the video-based HP test). The authors concluded that training was capable of improving HP performance, with a combination of on-road and video-based training resulting in the best improvements to HP performance.

Chapman, Underwood and Roberts (2002) examined the effects of training on the eye movements of two groups of novice drivers over their first year of qualified driving. The driver's eye movements were monitored while driving a specified route on public roads, as well as during a video HP test. One of the groups received training while the other did not. The training involved informing the drivers about how they visually searched driving scenes and then instructing and training them in the importance of scanning the visual scene in multiple locations in order to detect sources of potential danger. The training also involved HP videos that were stopped at critical moments and the participants were tasked with anticipating what happened next or answer a question regarding what had just happened. The training resulted in significant changes in the way the trained drivers scanned the visual scene during both on road driving and the video based hazard task. The trained participants reduced their fixation durations and increased their horizontal scanning of the visual scene, which is similar to the pattern displayed by older and more experienced drivers.

A large amount of work regarding HP testing has been carried out by research groups at the University of Massachusetts, Amherst. The research has focused on PC-based training programmes to improve HP. Pollatsek et al. (2006) used a PC-based training programme to see if they could improve the perception and awareness of inexperienced drivers to hazardous situations. The training programme involved providing the drivers with a top down views of hazardous scenarios that helped the drivers identify where potential risks were and what areas needed to be attended to. Learner drivers who had taken the training were compared to leaners that had not received any training. Their eye movements were recorded as they drove through 16 simulated scenarios (consisting of scenarios that were used in the training others that were not) in a driving simulator. The results revealed that the trained learner drivers were 22% more likely to look at areas containing information that had the potential to reduce the likelihood of a crash.

This is supported by research from Pradhan, Pollatsek, Knodler and Fisher (2009) who used a similar but further developed PC-based training course. The training consisted of presenting snapshots of driving scenarios in which the participants had to identify areas they would need to pay attention to if they were driving. The participants were also shown top down views of those same driving scenarios, which contained explanations of the hazardous elements of those scenarios. The purpose of this training was again to help drivers identify where potential risks were and what areas needed to be attended to. A further difference is that they did not test the effects of training in a simulator but in driving on public roads. A 16-mile route with 10 driving scenarios was selected for testing. Half of the scenarios were similar in concept to those used in training or had been photographed for the training session. The other half was different from the scenarios seen during training. They tested 18 to 21 year old drivers who had held a driving licence for at least 1 year. Their results showed that the drivers who underwent training looked at areas that contained information that could reduce the risk of a crash 64% of the time, compared to 37% of the time for the untrained drivers. These effects were also present for road situations that were significantly different from those shown in training.

McKenna and Crick (1994) noted that highly trained police officers displayed superior HP performance to experienced drivers. One major aspect of police driver training involves commentary driving. Commentary driving consists of a driver providing a verbal commentary that covers what they can see and what they are thinking. The police have long viewed commentary driving as a way to improve HP skills. Crundall, Andrews, van Loon and Chapman (2010) this with the use of a driving simulator. Leaner drivers between the ages of 17 and 25 were initially assessed as they drove through a simulated route containing 9 hazards. The participants were split into two groups, with one receiving commentary training and the other receiving no training. The commentary training consisted of a classroom based lesson and an on-road segment in which a driver instructor taught them how to provide a commentary. The participants then drove through the same simulated driving route as before. The drivers who had received commentary training were found to crash less, reduce speed sooner when approaching hazards and were quicker to apply brake pressure than the untrained drivers.

1.3.7 Age, Experience, Accident Liability and Hazard Perception

The above research demonstrates that there are differences in HP as a function of age, driving experience and crash liability. Although not all of the findings have been in agreement, research indicates that these differences are likely due to methodological issues, namely the content of the hazards used. Research has uncovered potential explanations as to why these different groups display differences in HP performance.

One recurring finding is that young novice drivers differ from older experienced drivers in how they visually scan hazards and driving scenes in general. Over 40 years ago, Mourant and Rockwell (1972) reported that young novice drivers tended to look straight ahead in order to maintain their lane position, at the expense of processing other areas of the driving scene. The young drivers demonstrated smaller horizontal scanning compared to the older and more experienced drivers. According to Mayhew, Simpson Williams and Ferguson (1998), novice drivers are less aware of hazards in the environment than older experienced drivers and as a result are slower to respond to hazards. They stated that this was likely to be because "their search and scan abilities are less developed than those of more experienced drivers. The novice driver may not have fully developed the use of peripheral vision and because they have a smaller range of scanning of the roadway than experienced drivers, they can fail to detect hazards in the environment" (Mayhew et al., 1998, p.4). In an analysis of the police reports on over 2000 crashes involving young drivers between the ages of 16 and 19, it turned out that 43% of the crashes were due to failing to scan the road for information that could reduce risk (McKnight and McKnight, 2003).

Pradhan et al. (2005) analysed the eye movements of drivers of different age groups (16-17 year old learners, 19-29 and 60-75 year old qualified drivers) as they drove through simulated driving scenarios that differed in critical areas that were necessary to scan in order to reduce risk. The results showed that the 16 and 17 year olds scanned the critical areas only 35% of the time compared to 50% and 66% for the drivers between the ages of 19 and 29, and 60 and 75 respectively. The younger drivers also responded is a less safe way to the risks through their driving behaviours in the simulator. Garay et al. (2004) used a driving simulator to compare HP performance between experienced and inexperienced drivers in daytime and night-time conditions. They found that inexperienced drivers were significantly less likely to scan the scene for risk during daytime and night-time conditions. Even during daytime conditions, inexperienced drivers scanned critical areas less than 50% of the time. This research indicates that the inferior HP abilities and increased crash liability of young novice drivers may be a result of their failure to scan critical areas of driving scenes.

Chapman and Underwood (1998) recorded the eye movements of novice and experienced drivers while they watched video hazards. They found that during hazardous events, eye movements were characterised by narrower visual search. The authors stated that the effect was similar to attention focusing. They also found differences in visual scanning between the novice and experienced drivers. Novice drivers had longer fixation durations. This was attributed to them having difficulty processing the visual scene, whereas experience allows for greater processing. This effect was especially prevalent during dangerous situations, indicating that the novice drivers did not have the experience to deal with these situations. Experienced drivers were found to fixate lower down as well as having less vertical variation in fixation locations. Differences in eye measures were also found as a function of the road environment. The least complex rural roads evoked the longest fixation durations and the shortest spread of search, whereas the most complex urban roads were associated with the shortest fixation durations and the largest spread of search. This again demonstrates how differences in the properties of hazards can affect responses.

Crundall, Chapman and Underwood (2002) also investigated the differences between eye movements in novice and experienced drivers when they viewed hazards, with a focus on peripheral vision. Participants were tasked with responding to hazards in videos while a secondary task required them to respond to peripheral stimuli. The results revealed that viewing hazards resulted in attention focusing, with resources being redirected from extra-foveal regions to the point of fixation. There were differences due to experience, with inexperienced drivers slower to respond to events in their peripheral vision. Experienced drivers were also found to have greater levels of horizontal search and a different time-course of attentional deployment, withholding attention from important stimuli until needing to fixate it foveally in order to devote full attentional resources to it. As noted previously, Crundall et al. (2012) found that inexperienced drivers differed from experienced in terms of the way they scanned hazards depending on the precursors to those hazards.

McKenna and Crick (1994) reported that hazards that required drivers to look far ahead were best at discriminating between experienced and inexperienced drivers. Mourant and Rockwell (1972) reported that novice drivers tended to look straight ahead in order to maintain their lane position, at the expense of processing other areas of the driving scene. They suggested that inexperienced drivers may struggle to identify distant hazards is due to an inability to coordinate visual input with the physical movements required to control the car. Based on this, McKenna and Crick argued that inexperienced drivers might in effect be struggling from a dual-task performance deficit as both HP and vehicle control are demanding tasks for them. Experienced drivers on the other hand have automated vehicle control and therefore there is no dual-task deficit. This would have a limited effect in video-based HP tests as vehicle control is not required but it would effect driving on public roads or in a simulator. This effect would be seen in simulated HP tests though. Endsley (1995) proposed that low SA and high workload could lead to errors in performance. It is therefore possible that the high levels of workload for inexperienced drivers combined with low levels of SA result in inferior task performance for HP. As SA develops with experience (Endsley, 1995), young and inexperienced drivers can be expected to have lower levels of SA for driving than older and more experienced drivers.

Wallis and Horswill (2007) proposed that differences in HP performance due to experience are a result of differences in signal detection between these groups. They carried out a fuzzy signal detection analysis of two models describing individual differences in HP. Fuzzy signal detection theory was proposed by Parasurman, Masalonis and Hancock (2000), and combines signal detection theory with fuzzy logic. The aim of fuzzy signal detection is to provide a method of analysing decision-making. One of the models was based upon the idea that novice drivers are worse than experienced drivers at discriminating more hazardous and less hazardous situations. The second model posited that novice drivers require a higher threshold of danger before becoming aware that a situation is hazardous. Inexperienced (4 years or less driving) and experienced drivers (10 years or more driving) were tested. The inexperienced drivers were split into 2 groups, one that received HP training and one that did not. The HP training consisted of 17 minutes of hazardous driving scenes accompanied by commentary from a professional driving instructor, who provided tools for identifying hazard cues and anticipating potentially hazardous situations (Raikos, 2003).

The testing consisted of a number of video-based HP testing measures in which the participants were told to respond to traffic conflicts, which were defined as a situation in which a collision or near collision would happen unless avoiding action is taken. There was a HP test in which the participants had to respond when they anticipated that a potential traffic conflict might occur during 20 traffic scene videos. There was also a hazard-rating task consisting of 23 traffic scenes that were blacked out at critical points. The participants then had to provide a score regarding the potential that a traffic conflict might occur within the next 5 seconds. Experienced and trained drivers were found to respond more often and faster than the novice drivers. The untrained novice drivers were found to be less likely and slower to respond to the less hazardous scenes in the tests. The models were then tested to see if they could account for the differences in performance between the different groups. The results of the analysis revealed that the second model was capable of this, but the first model was not, suggesting that differences in HP are related to novice drivers require a higher threshold of danger to perceive a situation as being hazardous. This is supported by research that has shown that young novice drivers have a lower perception of risk than older and more experienced driver (Matthews and Moran, 1986; Rhodes and Pivik, 2010). The results also provide further evidence that HP is a trainable skill, with trained novices performing in a similar manner to experienced drivers. Not only that but the method of the training, getting novices to anticipate cues for potential hazards, further indicates the importance of detecting hazard cues for successful HP performance.

Another explanation for differences in HP performance between different groups of drivers focuses on individual differences in the perception of hazards and risk. Young and inexperienced drivers have consistently been found to overestimate their driving abilities and underestimate the level of risk while driving (Bragg and Finn, 1982; Matthews and Moran, 1986), as well as being significantly more likely to engage in dangerous driving behaviours and be involved in crashes. This may impact on their ability to perceive hazards as young novice drivers fail to perceive the risk of traffic situations. If a novice driver perceives a low level of risk then they will be less likely to respond to hazards in an appropriate way. Experienced drivers have been found to perceive driving situations holistically, whereas novice drivers perceive in a piecemeal way, devoid of context (Milech et al. 1989). Benda and Hoyos (1983) reported that novice drivers assess hazards on the basis of a single property, meaning that all situations that share a common property (e.g. an icy road surface) are perceived as containing the same level of danger. Underwood et al. (2005) reported that experienced drivers perceived video clips containing driving hazards to be more hazardous that inexperienced drivers. Liu et al. (2009) argued that the results of their study might have been in part due to experienced riders possessing greater levels of risk perception. They found that the novice riders who were inexperienced drivers (all of whom were between the ages of 19 and 21) were more confident in their HP abilities than the novice riders who were experienced drivers. This overconfidence in their abilities may have led to them underestimating the risk in the riding scenarios.

1.3.8 UK Hazard Perception Test

One factor that has emerged in driving research is the fact that young novice drivers are significantly more likely than older and more experienced drivers to be involved in road accidents and to die or be seriously injured on the roads. An OECD report (ECMT, 2006) revealed that globally, 16-24 year old drivers are significantly over-represented in crash and fatality statistics. Despite the age group representing only 10% of the populations of OECD countries, young drivers represent 27% of all drivers killed in OECD countries. The report also found that the death rate for 18-24 year olds is about double that of older drivers and that motor vehicle crashes are the single greatest killer of 15-24 year olds in OECD countries (ECMT, 2006). It is possible that one of the reasons for this is that the HP abilities of young and inexperienced drivers are inferior to those of older and more experienced drivers.

One of the reasons put forward to explain the high accident rates for young novice drivers is that they have underdeveloped HP abilities. As mentioned previously, research has indicated a link between slow HP reaction times and a higher likelihood of a driver being involved in crashes (McKenna and Horswill, 1999; Pelz and Krupat, 1974; Quimby and Watts, 1981; Quimby et al., 1986;). Meanwhile a study by Drummond (2000) linked slow HP reactions times to an increased likelihood of being involved in a fatal crash within the first year of driving. Quimby et al. (1986) found that longer HP reactions times were associated with higher crash rates after controlling for age, driving experience and simple reaction times. Research has found that inexperienced and untrained drivers have slower HP reaction times (McKenna and Crick, 1994; McGowan and Banbury, 2004; Wallis and Horswill, 2007), while young drivers have slower HP reaction times than middle-aged drivers (Quimby and Watts, 1981).

As a result of findings such as these, as well as findings showing that HP is a trainable skill (McKenna and Crick, 1994; Mills et al. 1998), the UK Government decided to implement a HP test that all learner drivers would have to pass in order to gain a full UK driving licence (first implemented in 2002). The logic behind the implementation of the HP test was that learner drivers who fail to respond quickly enough to driving hazards in video clips might fail to respond quickly enough to on-road hazards. This could potentially lead to these drivers being involved in more crashes if given a full driving licence. The finding that HP performance can be improved through training indicated that a proper training and testing system could improve the HP of drivers. For example, HP training and testing could potentially lead to learner drivers and their instructors focusing more on HP while driving, which in turn could reduce accident rates in novice drivers. Although the test was implemented in 2002, the history of the UK HP test goes back a few years before this.

The current UK HP test used by the Driver and Vehicle Standards Agency (DVSA) was designed and validated by Grayson and Sexton (final report published in 2002), with assistance from the DSA (Driving Standards Agency- now part of the DVSA) and the Transport Research Laboratory (TRL). The aim of the research project was to create a test that could reliably distinguish between the HP abilities of experienced and inexperienced drivers, in turn making sure that only people with sufficient HP abilities could pass the test and receive a driving licence (with the aim of improving road safety). The design was heavily influenced by the video-based HP methodology pioneered by McKenna and Crick (1994) and Grayson and Sexton defined hazards on the basis of McKenna and Crick's definition of a hazardous situation. On the basis of this definition of a hazardous situation, they created 96 written scenarios (scripts) to guide the filming of HP clips. These scenarios were based on 6 environmental scenarios (rural lanes, suburban roads, urban roads, residential areas, single carriageways and dual carriageways) and 3 hazard categories (entry of another object into the path of the vehicle, unexpected actions of the vehicle in front and hazardous action of oncoming traffic). The scripts were then used to design and film the HP clips, which were filmed from the bonnet of a car driving on UK public roads. The clips were either staged or filmed on an opportunistic basis.

1057 inexperienced or experienced drivers took part in the validation test of the HP clips and testing procedure. Four tests were produced (each having a combination of different HP clips that were the best etc.). Participants were presented with the HP clips and had to respond to a hazard with a button press (some clips had more than one hazard). Responses to the hazards were scored on the basis of hazard windows. These hazard windows defined the time points at which making a button press would result in a score being given, with button presses outside of this window not be given a score. The start of the hazard window marked the onset of the hazard, the time at which the hazard became perceptible and started to develop. The end of the window marked the occurrence of the hazard, the point at which avoiding action would now need to be taken.

Grayson and Sexton (2002) reported that each of the four tests produced high levels of reliability (vital for psychometric testing), and that three were able to distinguish between experienced and inexperienced drivers. However, a number of studies (which used the HP clips designed by Grayson and Sexton) found that these tests were unable to demonstrate an association with accident liability. Chapman and Underwood (1998) failed to find differences in the scores of novice and inexperienced drivers using the HP clips designed by Grayson and Sexton. They went on to suggest that differences in hazard types were in fact more important than the experience levels of the drivers. Field, Groeger and Hammond (1997) reported no correlation between hazard scores and any measure of accident involvement. Grayson and Maycock (1997) also failed to find any association between hazard scores and involvement in accidents.

This led to a further program of work with the TRL, with the aim of producing HP tests that would correlate with accident liability and driver experience. The HP items that were better able to discriminate experience levels were analysed in order to understand what properties of the items allowed for this discrimination. The 'best' clips required drivers to thoroughly scan the scene as well as anticipate potentially hazardous events (which is similar to the findings of other HP testing and research). These clips were capable of discriminating driver experience as well as having “some association with retrospective accident liability” (Grayson and Sexton (2002), p.20). This information was then used to create a 'blueprint' that defined the types of HP items that were required to develop successful HP clips.

From this blueprint, new HP clips were filmed and tested on drivers with varying levels of experience. Grayson and Sexton concluded that the HP items possessed sufficient reliability and validity to create a new set of HP tests. Finally, Grayson and Sexton tested whether or not the new HP training could improve scores on the HP test. The DSA created a training package that learner drivers, ready to take their theory test, were taught from. These learners then took the newly designed HP tests. The results revealed that the use of the training package led to learner drivers achieving similar HP average scores as those of experienced drivers. This HP trial was then developed and expanded into the full UK HP test as it is administered today. Only items that were able to discriminate between learners, inexperienced and experienced drivers were included.

1.3.9 Effect of The UK HP Test

Given the reasons for introducing the HP in the UK, the most important factor is whether or not the test has reduced accident rates among newly qualified drivers. Wells et al. (2008) evaluated the impact that the HP test had on driving in the UK. They carried out a longitudinal study on newly qualified drivers who had not taken the HP test and those who had taken the recently introduced HP test. They used questionnaires to gather information on any accidents, driving offences etc. They found that the HP test only had benefits for specific driving situations. They reported that the HP test was associated with reductions in subsequent accident liability in the first year of driving for some specific accidents, with the size if the effect dependent on the accident type. For reported non low-speed accidents on a public road where the driver accepted some blame, the accident liability of those who had taken the hazard perception test was significantly lower than those who had not. Findings such as this do provide some support for the claim by Grayson and Sexton (2002) that the revised HP clips are associated with accident liability. Wells et al. (2008) also reported that drivers who had taken the HP test committed fewer driving faults during their practical tests than those who had not taken the HP test. In particular, the drivers who had taken the HP test made fewer awareness and anticipation faults. There was also a predictive relationship between HP scores and the number of reported accidents during the first year of driving.

HP is now firmly established as a critical part of driver training and testing and one that continues to develop. The work in this thesis is designed to help us understand more about the HP test as currently used in the UK, to explore some of the issues in developing the HP test into a full simulated environment and explore some of the neural correlates of driving in hazardous situations.

Chapter 2 - UK Hazard Perception Test

2.1 Abstract

There has yet to be a full and systematic investigation into the results of the UK Hazard Perception (HP) test. In particular, there has not yet been an examination into the HP videos themselves, namely whether or not there are particular properties of the hazard videos which make them better at discriminating between drivers with and without the required levels of HP. In order to achieve this, a multidimensional scaling (MDS) was carried out on a very large set of data from the UK HP test. The MDS examined how test candidates performed on the individual hazards videos within the HP test in order to determine whether or not there was any underlying structure in the data (i.e. if certain hazard types are best at predicting passes or fails on the test). The results of the MDS revealed that there was no underlying structure in the data that would point to particular types of hazards predicting performance on the HP test. Explanations for this finding are discussed in detail. Possibilities for further research and alternative approaches that can extend and improve upon the findings of this study are also discussed.

2.2 Introduction

2.2.1 Hazard Properties

As detailed in Chapter 1, research has indicated that the content of HP clips affects their ability to discriminate between drivers. In particular, it appears that the precursors (or cues) to the hazards, the properties of the hazard and the level of hazard anticipation required interact to affect HP differences in experienced and novice drivers.

As mentioned previously, a large number of studies have found that the properties of hazards affect their ability to discriminate between novice and experienced drivers. A number of studies have indicated that the key discriminatory factor of a hazard relates to whether or not there are any precursors or cues to the hazards (Garay-Vega et al. 2007; Sahar et al. 2010; Crundall et al. 2012). Given the research that has also shown that novice drivers differ in the way that they scan the road (Mourant and Rockwell, 1972; Chapman and Underwood, 1998; Crundall, Chapman and Underwood, 2002) and are much less likely to scan critical areas than experienced drivers (Pradhan et al. 2005; Garay et al. 2004), it is possible that novice drivers failure to scan important areas of the visual scene leads to them missing precursors to hazards. This may in turn be one of the reasons their HP abilities are lower and they are involved in more crashes. This is supported by the work of Crundall et al. (2012) who found differences in terms of how experienced and inexperienced drivers scanned hazards depending on the precursors to those hazards. Wells et al. (2008) reported that newly qualified drivers who were less likely to be involved in crashes had better skills of anticipation, perception and observation.

Research has also shown that training drivers to look for and anticipate cues to hazards results in improved HP performance (Raikos, 2003). Some researchers have also linked precursors to SA (Underwood et al. 2010; Crundall et al. 2012), with a particular focus on level 3 (anticipation) of Endsley's (1995) model of SA. This also fits with research that has found that training drivers in hazard anticipation can improve HP performance (McKenna, 2006). This may be related to the idea that SA is a trainable skill that develops with experience, and training provides that experience. However, simply placing cues to hazards into videos or simulations does not result in them discriminating between drivers of different experience or training levels. Garay-Vega et al. (2007) used precursors in their hazards yet they were unable to differentiate between experienced and inexperienced drivers. This is perhaps because the precursors they used were too obvious and easy to spot.

Overall this research demonstrates that the structure of hazards used in a HP test affects their ability to discriminate between experienced and inexperienced drivers. Given that the UK HP test was designed to be able to discriminate between drivers on the basis of their levels of driving experience, these findings are important as they indicate what factors best discriminate between experienced and inexperienced drivers, which in turn is designed to be the measure of discriminating between drivers that have the required levels of HP to pass the test and those that do not.

2.2.2 UK Hazard Perception Test

In its current form the UK HP test consists of 14 hazard video clips of on-road driving. During the time this research was being conducted the UK HP test consisted of 14 hazard video clips of on-road driving. The videos are taken from the perspective of the driver. The videos are presented on computers and the test candidates respond to them using a mouse. 13 of the hazard clips contain 1 scoring hazard while 1 of the clips contains 2 scoring hazards. Each hazard is scored on the basis of response times in relation to the hazard windows. These timing windows were established by the DSA. They went through each video and assigned a time point that indicated the hazard onset and another time point that indicated that the hazard occurrence. Each window was divided up into five sections, meaning that if a test candidate responded during the hazard window they could score between 5 points and 1 point depending on how quickly they responded. As a result there are a total of 75 points available during the HP tests. In order to pass the HP portion of the theory test, the candidate must currently score at least 44 points. There were a total of 150 HP video clips in usage by the DSA. The video clips are arranged into 20 tests sets that each contain 14 hazard video clips (many clips appear in more than one set). Each set has been chosen to balance difficulty and to include a range of different road scenarios. The test sets are randomly assigned to test candidates. Before the test begins an introductory video is played that explains what the test candidate is required to do during the test. The UK HP test defines a hazard as “something that might make you, as the driver, slow down or change direction”.

A recent change to the UK HP test involves the introduction of animated video clips. As the video clips used in the HP test were filmed a number of years ago, they started to look dated as many of the models of cars in the video are rarely on the roads any more and the resolution of the videos were relatively poor in comparison to modern standard. As a result of this the DSA were concerned that the videos may start to look unrealistic in the eyes of the test candidates. Re-filming the clips was not practical due to the time and cost involved but most importantly it would not have been possible to produce exact copies of the original HP clips. This would have resulted in the need to validate the new clips, which would be time consuming and expensive in itself. Due to this an animation company was hired to produce exact computer generated (CG) copies of the HP clips. These CG clips were introduced into live tests in January 2015.

2.3 Study 1

So far there has not been a systematic study of data from the HP test carried out in order to explore if there are any underlying structures to performance on the HP test. Therefore the aim of Study 1 is to explore if there are any underlying factors that can account for performance on the HP test. In particular, does performance on certain HP test clips predict overall performance on the HP test (e.g. a pass or fail on the test)? This process will involve analysing data from the UK HP test that has been provided by the DVSA using a multidimensional scaling. A multidimensional scaling will be used as it provides a way of visualising what is a very large and complex data set. A multidimensional scaling provides the ability to visualise the similarity of the cases in a dataset (in this case the hazard videos in particular HP test sets). This in turn provides the ability to determine how individual hazards contribute to test candidates passing or failing the HP test. The results of this study may help inform the DVSA about which HP videos are best at discriminating between drivers who posses the appropriate HP skills and those who do not. This could in turn help the DVSA refine aspects of the HP test.

2.4 Methodology

2.4.1 Participants

The participants consisted of all the test candidates who took the UK theory-driving test during the month of February 2012. The decision was taken to only analyse data from test candidates who were taking the HP test for the first time in order to avoid any affects multiple attempts may have had on test performance. After the data was sorted and the required information was extracted the total number of participants was 26,452. A total of 32 test candidates did not provide a gender. A total of x passed the HP test, while x failed to pass the test. A more detailed breakdown of the participant information can be found in Table 2.1.

Table 2.1. Details of candidates included in the HP test analysis.

Test Set	Pass Or Fail	Candidates	Male	Female	None
HP0042	Pass	2888	1434	1454	0
HP0042	Fail	388	239	149	0
HP0043	Pass	2673	1353	1319	1
HP0043	Fail	554	314	240	0
HP0044	Pass	2972	1479	1491	2
HP0044	Fail	372	205	167	0
HP0045	Pass	3052	1533	1516	3
HP0045	Fail	337	201	136	0
HP0046	Pass	2963	1479	1484	0
HP0046	Fail	428	236	192	0
HP0048	Pass	2885	1455	1430	0
HP0048	Fail	326	195	131	0
HP0049	Pass	2830	1394	1434	2
HP0049	Fail	425	248	177	0
HP0050	Pass	2862	1474	1386	2
HP0050	Fail	454	266	188	0
HP0051	Pass	2904	1427	1477	0
HP0051	Fail	315	193	122	0
HP0052	Pass	2887	1447	1439	1
HP0052	Fail	403	224	179	0
HP0053	Pass	2962	1454	1507	3
HP0053	Fail	387	234	152	1
HP0054	Pass	2785	1394	1389	2
HP0054	Fail	541	301	240	0
HP0055	Pass	2827	1387	1438	2
HP0055	Fail	404	194	210	0
HP0056	Pass	2859	1425	1433	1
HP0056	Fail	326	205	121	0
HP0057	Pass	2814	1400	1413	2
HP0057	Fail	458	268	190	0
HP0058	Pass	2774	1375	1399	0
HP0058	Fail	514	297	217	0
HP0059	Pass	2947	1472	1471	4
HP0059	Fail	405	214	191	0
HP0060	Pass	2913	1467	1445	1
HP0060	Fail	472	270	202	0
HP0061	Pass	2940	1465	1473	2
HP0061	Fail	438	229	208	1
HP0062	Pass	2981	1531	1449	2
HP0062	Fail	380	213	167	0

2.4.2. Materials

Each HP test consists of a test set containing 14 HP videos. 13 of the video clips contain 1 score-able hazard and one of the clips contains 2 score-able hazards. A maximum of 5 points is awarded for each score-able hazard. Points awarded are determined on the basis of scoring windows, which define points in time in which the hazard becomes visible and when responding to the hazard would be too late to avoid having a collision. A button press in the first fifth of this time window provided the candidate with a score of 5, while scores in later windows receive lower scores, down to 1 for the last window. Any button responses made outside the scoring window receive no points. The HP test sets are assigned to test candidates at random. There are 20 HP test sets and 150 different HP videos in total (including 4 videos that contain 2 soring hazards), with some HP videos shared between test sets. The HP test data contained:

- Registration number (to identify the test candidate)
- Gender of test candidate
- Age of test candidate
- Test attempts
- The test set the candidate was assigned to
- The individual hazard videos the test candidates saw
- Scores for each hazard video
- Total score for the HP test
- What grade the test candidate received for the HP test

2.4.3 Design

The main analysis of the data involved an ALSCAL multi-dimensional scaling on the scores test candidates received for the HP video clips they saw. ALSCAL is a multidimensional scaling algorithm developed by Young, Takane and Lewyckyj (1978). ALSCAL is a metric and nonmetric multidimensional scaling method that allows for a number of individual difference options. ALSCAL uses the alternating least squares approach (Takane, Young and Lewyckyj, 1977). ALSCAL unfolds the data in order to construct a Euclidian space, which has points for each object. Rows and columns of the data are represented as points within that Euclidean space.

ALSCAL analyses the similarity of values within a proximity matrix in order to create a scaling solution that represents the similarity/dissimilarity of points within Euclidian space. If data points (in this case scores for particular HP test videos) are similar, then they are placed close together, while if the data points are dissimilar, the points are placed far apart. The ALSCAL goes through an iterative process of unfolding the data to produce the scaling solution. Each step involves making slight adjustments to the scaling solution by moving the data points in order to improve the reliability of the solution. The iterations are stopped when any further iteration results in an improvement in stress (i.e. the reliability of the scaling solution) of less than .001. The end result of the multidimensional scaling is that the data are placed into a plot in order to provide a visual representation of the scaling solution (i.e. how similar/dissimilar the data points are in the dimensions of Euclidian space). The analysis also outputs goodness-of-fit values (the proportion of variance accounted for by the multidimensional scaling) and stress values for the scaling solution.

The data were analysed in 2-dimensions of Euclidian space. This value was chosen, as this is the maximum number of dimensions possible to produce reliable results with the amount of data from each HP test set (i.e. the number of data points in Euclidean space). Separate scaling solutions were created for candidates who passed the HP test and those who failed the test.

2.4.4 Procedure

The first step taken was to place the data into an MS Access database as this enabled easy extraction of specific parts of the data set for analysis. From this database, 20 sets of data were extracted, one for each of the HP test sets. These contained the scores that each test candidate received for the 14 hazard videos that make up each test set. From these test sets, data was extracted for test candidates who were attempting the HP for the first time. This was done in order to avoid any practise effects that may have existed in candidates who had attempted the HP test multiple times. This allowed for a clean and direct comparison between test candidates without any confounding effects of multiple testing.

The resulting data sets were then analysed in SPSS using an ALSCAL multidimensional scaling in order to establish proximity values between the HP video clips contained within each of the 20 HP test sets. Separate proximity matrices were calculated for those who had passed the HP test and those that had failed the HP test. In total, 40 proximity matrices were created, 2 for each of the 20 HP test sets (one for the participants who had failed the HP test and another for those that had passed the HP test). After inspecting these 40 proximity matrices, it became apparent that including the video hazard clips that contained double hazards resulted in the scaling solutions being distorted by the presence of the double hazards (as the dimensions of the scaling solutions were highly skewed by the double hazards). As a result the multidimensional scaling was repeated, this time with the double hazard clips removed, leaving 13 scores for each test candidate in the analyses.

2.5 Results

Table 2.2 details the Kruskal stress test (goodness-of-fit) values and the proportion of variance explained by the multidimensional scaling (r-squared) for each of the HP test sets, for those who passed or failed the HP test. The R squared values demonstrate that the multidimensional scaling solutions were successful for most of the HP test sets. R squared values range from 0.74 to 0.98, indicating that the scaling solutions were successful at accounting for the variance in the data.

In order to compare how well the scaling solutions fit the data for those that passed the test and those that failed, t-tests were used to compare the r-squared and stress values for the scaling solutions. The t-test comparing the R-squared values revealed no significant difference between the passes and the fails [$t(38)=1.513$, $F=0.150$, $p=.701$]. The t-test for the stress test revealed no significant difference between the passes and fails [$t(38)=0.508$, $F=0.182$, $p=.615$]. These findings show that the modelling solutions for passes and fails were not significantly different from one another.

Table 2.2. Stress values and effect sizes for the multidimensional scaling on the HP test data.

Test Set	Pass Or Fail	Stress	R^2
HP0042	Pass	0.174	0.86
HP0042	Fail	0.141	0.9
HP0043	Pass	0.182	0.88
HP0043	Fail	0.164	0.87
HP0044	Pass	0.08	0.98
HP0044	Fail	0.17	0.93
HP0045	Pass	0.166	0.88
HP0045	Fail	0.162	0.87
HP0046	Pass	0.131	0.92
HP0046	Fail	0.149	0.89
HP0048	Pass	0.095	0.98
HP0048	Fail	0.177	0.84
HP0049	Pass	0.123	0.93
HP0049	Fail	0.128	0.94
HP0050	Pass	0.126	0.95
HP0050	Fail	0.175	0.84
HP0051	Pass	0.176	0.87
HP0051	Fail	0.139	0.91
HP0052	Pass	0.157	0.92
HP0052	Fail	0.153	0.89
HP0053	Pass	0.163	0.9
HP0053	Fail	0.211	0.82
HP0054	Pass	0.131	0.91
HP0054	Fail	0.149	0.9
HP0055	Pass	0.261	0.74
HP0055	Fail	0.215	0.77
HP0056	Pass	0.11	0.96
HP0056	Fail	0.097	0.96
HP0057	Pass	0.12	0.96
HP0057	Fail	0.091	0.97
HP0058	Pass	0.15	0.91
HP0058	Fail	0.194	0.82
HP0059	Pass	0.151	0.93
HP0059	Fail	0.167	0.87
HP0060	Pass	0.1	0.96
HP0060	Fail	0.1	0.95
HP0061	Pass	0.117	0.96
HP0061	Fail	0.09	0.96
HP0062	Pass	0.178	0.93
HP0062	Fail	0.143	0.91

A detailed breakdown of the mean scores for each HP video in each HP test set, for both passes and fails, can be found in Appendix 2.1.

In order to provide a clearer visual representation of the scaling solutions and help interpret the dimensions in those solutions, scatter plots were created using the proximity values produced by the multidimensional scaling carried out in SPSS. The individual hazards in each test set were assigned to specific categories on the basis of hazard classifications provided by the DVSA. The hazard classifications provided detailed information on the events that too place in each HP test video. This included factors such as the road environment (e.g. residential area), the weather and what was responsible for the hazard (e.g. a pedestrian stepping into the road). The points for each individual hazard in the scatter plots were assigned a specific colour depending on the primary cause of the hazard and a particular shape depending on the road environment in which the hazard took place. A total of seven hazard categories and five environment categories were created.

The hazard categories were: vehicles emerging from junctions/side-roads (red), oncoming vehicles (green), obstacles in the road (yellow), cyclists (gold), pedestrians (blue), animals in the road (purple) and erratic/dangerous driving (black). The environment categories used were: urban (diamond), residential (circle), rural (square), dual-carriageway (triangle) and commercial/industrial (dash).

The scaling solutions in Figure 2.1 to Figure 2.4 detail the scaling solutions for two of the HP test sets (HP42 and HP43), for both the passes and the fails, with the individual hazards having been assigned to the specific colour and shape categories listed above. The remaining figures for the scaling solutions for the rest of the HP tests sets (for both passes and fails) can be found in Appendix 2.2.

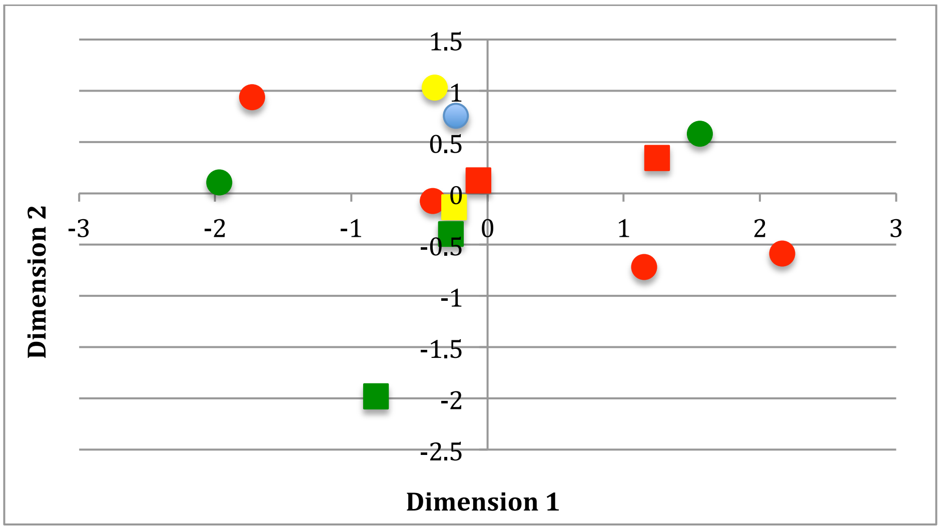


Figure 2.1. Figure displaying the multidimensional scaling solutions for the passes for test set HP042. Each data point represents one hazard video used in the test set, with colours and shapes corresponding to the hazard classifications.

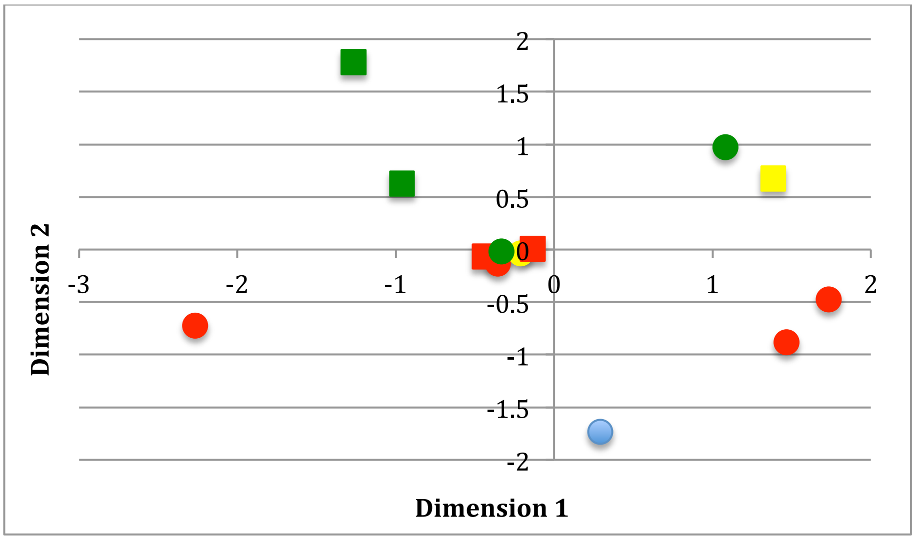


Figure 2.2. Figure displaying the multidimensional scaling solutions for the fails for test set HP042. Each data point represents one hazard video used in the test set, with colours and shapes corresponding to the hazard classifications.

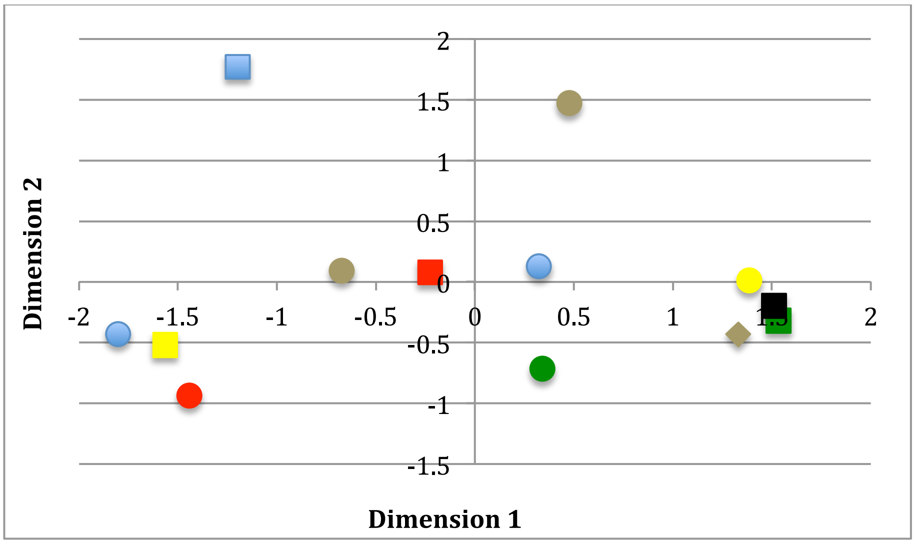


Figure 2.3. Figure displaying the multidimensional scaling solutions for the passes for test set HP043. Each data point represents one hazard video used in the test set, with colours and shapes corresponding to the hazard classifications.

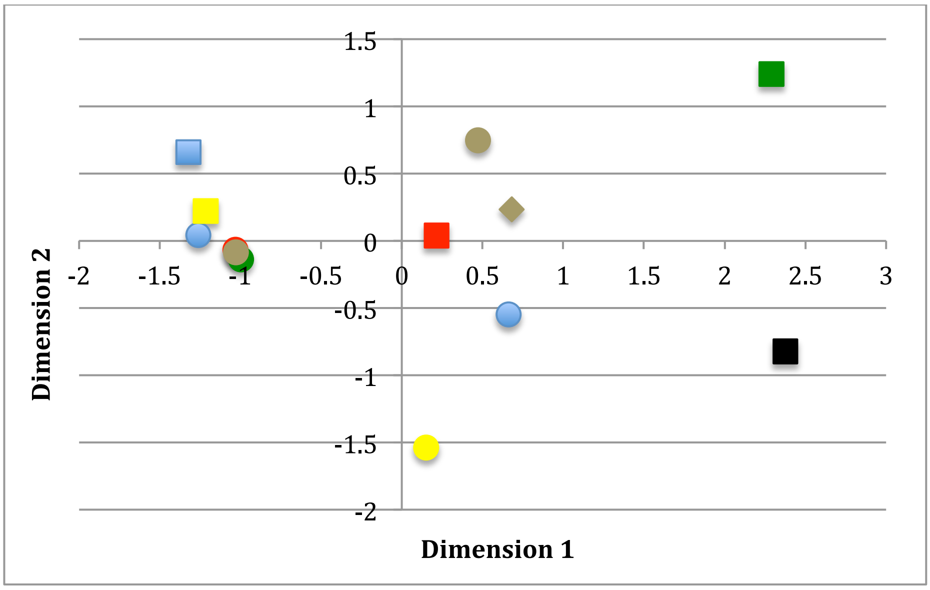


Figure 2.4. Figure displaying the multidimensional scaling solutions for the fails for test set HP043. Each data point represents one hazard video used in the test set, with colours and shapes corresponding to the hazard classifications.

2.6 Discussion

The aim of study 2 was to explore if there are any underlying factors that contribute to performance on the UK HP and to see of these factors could be used to help improve the HP test discriminate between drivers with and without the required level of HP ability. Unfortunately no evidence of an underlying structure was found through studying the scaling solutions for those who passed or failed the HP test. When inspecting the scaling solutions it becomes clear that they are rather idiosyncratic, with no clear pattern emerging across them. Scaling solutions for some of the test sets (for both passes and fails) indicate that the underlying structure is related to the 'erratic driving' hazards (e.g. HP61), whereas others indicate that hazards caused by obstacles in the road (e.g. HP44). This inconsistency is present for scaling solutions for both passes and fails, not only between the different HP test sets but also within them. For example, in the scaling solution for HP55 for passes, the pedestrian-residential hazards that are very similar in nature are placed in very different locations. There is also no clear pattern to bee seen when comparing the pass and fail scaling solutions for the same HP test set, with some matching each other relatively closely (e.g. HP57) and others differing significantly (e.g. HP52). These inconsistencies also apply to the environments in which the hazards took place.

One conclusion to take from the data is that the individual hazards and the hazard tests sets are fairly well matched with one another. It is possible that, overall, the individual hazards are measuring the same abilities and processes in the test candidates, while the test sets have a good balance of different hazards within them. It is also possible that in order to uncover any underlying structure in the HP test data, it will ne necessary to look at more subtle differences in the content of each of the hazard videos used in the UK HP test. These could include looking at the precursors to the scoring hazards and classifying the hazards on the basis (or at least in part) of the precursors. As mentioned in detail in chapter 1, previous research has detailed how novice drivers differ from experienced drivers in terms of how they detect and respond to certain types of hazards. The study by Crundall et al. (2012) details that certain hazards are able to distinguish between experienced and inexperienced drivers due to the precursors to those hazards and the relationship between the hazard and the precursor. The precursor to the hazard was found to affect how drivers detected and then responded to the hazards. For example, learner drivers were found to be particularly likely to miss hazards obscured by the environment (e.g. pedestrians stepping out from behind cars). Overall this suggests that investigating the underlying structure of a hazard may reveal an underlying structure within the HP test data.

Following on from the work of Crundall et al (2012), each one of the video hazards used in the UK HP test could be classed on the basis of the hazard precursors, the hazards and the relation between the two. Crundall et al (2012) defined hazard precursors on the basis of whether they were directly or indirectly related to the hazard. This type of precursor classification could be applied to each of the videos in the UK HP test and data from the test could then be analysed using a MDS that takes these classifications into account. Research from Shahar et al. (2010) has revealed that sudden onset hazards that lack adequate hazard precursors affect the ability of hazards to discriminate between experienced and inexperienced drivers. It is possible that a number of hazards lacking proper precursors are present in the UK HP test. Analysing differences in performance between hazards with and without of adequate hazard onsets could also help to provide further information on what types of hazards are best at discriminating between drivers with and without the necessary level of HP ability. Combining all of these factors together with the research already carried out in this study could help to establish if there is an underlying structure in the HP test and what that structure is. This in turn could be used by the DVSA to improve the UK HP test by altering which hazards are used in the test as well as helping them to design new hazards that could be used in the test.

A limitation of this study was the inability to include the double hazard videos in the MDS analysis. A way to improve this study would be to include the double hazards in the MDS. In order to do this is would be necessary to parse the scores for the double hazards so that the individual scores for each of the two hazards present in the double hazard videos could be ascertained. As mentioned in the methodology section, simply analysing the double hazards without parsing them results in the dimensions of the MDS being dominated by the double hazards (which removes important information about the relationships between the single hazards in each clip group). The failure to analyse the double hazards separately in this study was due to a technical issue that could not be overcome. However, there may be a way to parse the double hazards and then analyse them separately.

Due to the HP test being organised into a set of hazard clip groups, it was necessary to carry out individual MDS for each of the clip sets. A drawback to this method is that the similarity/dissimilarity of the hazards could only be determined for hazards that were in the same clip group. This is further complicated by the fact that certain clips are shared between multiple clip groups. Meanwhile each double hazard is shared between multiple clip groups due to the DVSA only having several double hazard videos at its disposal. This means that one possible explanation for the failure to find an underlying structure in the clip groups is that most of the variance in the data is accounted for by the allocation of the hazards to the clip groups themselves. It is possible that by analysing all of the hazards used in the HP test in a single analysis, an underlying structure in the hazards may emerge. The best way to determine if there is an underlying structure in the HP test would be to have a large number of learner drivers complete a HP test in which they watch and respond to each of the hazards currently used in the HP test. This would allow for a single analysis of all hazard videos rather than analysing the videos by clip group. However, due to the sensitivity of the HP test (as well as other logistical issues), the DVSA would likely not allow this. As a result of this, any analysis of the HP data will be hamstrung by the fact that the HP videos will need to be analysed by clip group.

A significant decision made in this study was to only analyse data from test candidates who were taking the HP for the first time. This decision was taken in order to avoid any practice effects that may emerge after multiple test attempts. It would however be interesting to carry out further analyses with the test candidates who were on the second or greater attempt at the HP test. Such an analysis could help to determine if there are certain types of hazards that best discriminate between those that pass and those that fail the HP test. A future analysis could focus taking data from test candidates who fail the test on one or more occasions.

A major way to extend the research in this study would be to take data the UK HP test and then compare that to data from the practical driving tests (i.e. the on-road test conducted by a driving examiner). By comparing how test candidates perform in the HP and practical tests, it may be possible to see if performance on the HP test is predictive of practical test performance (and driving ability on the road). Given that the fundamental aim of the UK HP test is to ensure that only test candidates with adequate levels of HP ability are able to take the practical driving test and be given the opportunity to gain a full driving licence, it is important to determine if the HP test is having the desired effect on real-world driving. Practical test performance from candidates who passed the HP test (measured in terms of whether the test candidate passed or failed the test, how many major and minor faults they received and what those faults were attributed to) could be compared with their HP test performance (what overall score they received, what scores they received for individual hazards and what hazard videos were in their test). This analysis could determine if performance on the HP test (and performance on particular hazard types) is predictive on driving behaviours. In turn this could help inform the DVSA if there are particular types of hazards that are best at predicting whether or not test candidates will pass or fail their test and if they fail their test, why they fail.

Chapter 3 - Behavioural and Physiological Differences Between a Hazard Perception Test and Simulated Driving

3.1 Abstract

Driving collision statistics have revealed that young novice drivers are significantly more likely to be killed or injured while driving on the roads than older and more experienced drivers. One of the reasons put forward to explain this effect is that young and inexperienced drivers have insufficient hazard perception (HP) abilities. This study examined differences in HP performance and physiological responses between video and simulated HP tests, with the aim of determining if simulated hazards may provide a more realistic of test of HP abilities than the video hazards which have predominantly been used to test HP abilities. The results of the study revealed significant differences in the behavioural and psychophysiological responses between the hazard videos and the simulated hazard scenarios. Overall, the findings demonstrate that hazard videos may not provide a representative test of the HP abilities used on the road, as well as demonstrating the technical difficulties associated with creating high quality simulated hazard scenarios. The implications for HP testing are discussed in detail.

3.2 Introduction

3.2.1 General Introduction

Study 2 involves utilising a driving simulator to further investigate HP and HP testing. As mentioned in Chapter 1, driving research is increasingly utilising driving simulators to explore driving abilities. There is also an increasing amount of research that is using driving simulators to investigate HP. However there is considerable variability in the way this has been done and little of it has been related directly to traditional video-based hazard perception testing. Bringing the interactivity and realism of driving simulation into HP testing may allow for improvements to HP testing by making it more comparable to HP while driving on the road. In order to explore this possibility, Chapter 3 will involve comparing performance on HP videos provided by the DSA (which are representative of the HP videos used in the UK HP test) with simulated versions of those clips created in a driving simulator. Behavioural and physiological responses to the video and simulated hazards will be compared in order to establish if simulated hazards result in significantly different responses compared to video hazards and if these responses suggestive that simulated hazards evoke responses more like those of hazards in real-world driving.

3.2.2 Using Driving Simulation to Test Hazard Perception

Chapter 1 also detailed a number of studies that used driving simulators to investigate HP (e.g. Crundall et al., 2012, Garay-Vega et al., 2007; Liu et al., 2009; Shahar et al., 2010). Driving simulators provide a number of advantages for driving research. One of the major advantages they provide is interactivity. Unlike video-based driving research, a driving simulator allows participants to 'drive' a simulated vehicle, making use of the controls of a car (steering, accelerator, brake, clutch etc.). If driving research wants as to test aspects of driving then allowing participants to properly interact in an experiment is surely beneficial. As a result of the interactivity, driving simulators can provide a high level of realism. This extra realism can take the form of the car controls, sitting in a real car while completing an experiment as well as the increased visual realism driving simulators can provide through large projection screens (with a more representative field of view than a TV screen for example) and realistic computer graphics.

The interactivity of driving simulators also allows experimenters to collect more complex behavioural responses such as speed, steering, acceleration, speed, lane position and distance to vehicles in front. With regards to HP, this extra data can provide information on preparatory behaviours to hazards (changing lane to avoid a hazard) and emergency manoeuvres used to avoid hazards (e.g. emergency braking). Experimenters can use speed, braking, steering and lane position data on the approach to a hazard to provide more complex behavioural information that can show that a participant has detected a hazard and they have selected certain behaviours to avoid it.

Driving simulators with a large field of view provide the advantage of being able to simulate hazardous events that occur to the side of or behind a car, as opposed to every hazard being presented from in front as is the case for video hazards presented on a TV. Research has revealed that certain types of visual cues are extremely important for car handling, for example the suggestion that drivers need to fixate the tangent point when they steer around corners (Land & Lee, 1994). The requirement to attend to these cues when controlling a car will interrupt the visual search for hazards (Schieber, Schlorholtz, & McCall, 2009). This process cannot be replicated by video-based hazards. Using simulation can therefore result in a more realistic level of demand on the visual system compared to watching video-based hazards.

Driving simulators also provide a number of advantages over on-road experimentation. Testing driving abilities on the road has safety issues as well as making it difficult to control experimental variables. There are also ethical issues. Research has shown that driving simulators provides advantages over on-road driving studies in terms of experimental control, safety and cost (Reed and Green, 1999). Nonetheless there are considerable problems brought about by the use of simulated hazards rather than video-recorded ones. A key disadvantage that driving simulators have compared to video hazards is that it is much harder to strictly control the experience that participants have during a driving simulation. For example, whereas viewers will experience hazards in the exact same way when watching a video, different driving styles (speed, lane position etc.) in a simulator may affect that way that hazards are interacted with - a hazard approached slowly may be far less dangerous and easier to anticipate than one approached rapidly. While this can be seen as a benefit to simulated hazards as it is more representative of real world driving, it also creates major difficulties when it comes to creating a standardised test. This makes controlling variables within driving simulation more difficult. There is also the issue that testing people in driving simulators can lead to simulator sickness.

An important question is whether or not driving simulators provide a reliable and valid alternative to video-based and on-road driving experiments. There is some evidence from driving research that driving simulators can be a valid proxy of on-road driving. Konstantopoulos, Crundall, and Chapman (2010) found that novice drivers differed from experienced drivers in the way that they visually scanned the road environment in a driving simulator, these differences were similar to those displayed between novice and experienced drivers when driving on the road. Shectman et al. (2009) directly compared driving performance on the road and in a driving simulator by looking at errors made by drivers when turning left or right at junctions. Their results revealed that there were no differences in performance between the on-road and simulator conditions. For a review comparing HP on the road, in simulators and using traditional video-based methods, see Underwood et al. (2011).

3.2.3 Physiology of Driving

One way to explore if simulated driving can provide a more realistic test of driving skills is through the recording of physiological measures. The fundamental assumption of using physiological measurements is that all aspects of human behavior evoke physiological reactions, with these physiological reactions being involuntary and uncontrollable (Bouscein & Backs, 2008). As a result, physiological recordings can provide a valid representation of a person's mental state, which in turn provides a means for studying the underlying cognition processes of various behaviors (including driving). Physiological measures are often recorded to assess the level of arousal of an individual. The level of arousal is measured along a continuum, ranging from low to high. For example, low arousal would be present during sleep while high levels of arousal would be present during physical exertion. Arousal operates via the sympathetic autonomic nervous system and takes the form of a variety of physiological reactions including an elevated heart rate, increased respiration rate, increased skin conductance (electrodermal dermal response) and pupil dilation. Arousal can also vary due to emotional reasons, for example as a result of stress. Importantly, arousal is also affected by cognitive changes, such as mental workload. As a result of this, physiological recordings can provide a measure of cognitive factors during tasks such as driving.

Two commonly used measures of physiological changes are heart rate and skin conductance. Heart rate is the speed at which the heart beats, usually measured in the number of beats per minute (bpm), as well as in terms of the change in heart rate and the variability of the heart rate. Changes in heart rate can be caused by factors such as exercise, stress and anxiety. As a result of this, heart rate can be used as a measure physical and psychological arousal. Skin conductance is a method of measuring the electrical conductance of the skin, which varies depending on the state of the sweat glands. Much like heart rate, sweating is controlled by the sympathetic nervous system; meaning that skin conductance is used as a measure of physiological and psychological arousal. When the sympathetic autonomic nervous system is aroused, the activity of the sweat glands increases, which in turn increases skin conductance. Skin conductance can be measured in terms of its mean, total or in terms of the number of specific electrodermal responses (EDRs) recorded over a period of time.

Research has revealed that physical and mental workload impact heart rate (de Waard and Brookhuis, 1991) as well as skin conductance (Boucsein, 1992). Mehler, Reimer and Coughlin (2012) found that skin conductance increased with increasing task demands, meaning that skin conductance can be used for discriminating between different task difficulties. As a result, these physiological measures can be used to measure changes in mental workload during psychological tasks.

Research has revealed that heart rate increases during effortful working periods of simulated work such as driving (Dijksterhuis, Brookhuis, and De Waard, 2011) and flying (De Rivecourt et al., 2008), real life work such as on-road driving (de Waard and Brookhuis, 1991) and during laboratory tasks such as working memory task (Backs and Seljos, 1994). Brookhuis and de Waard (2011) have advocated the use of physiological measures in combination with driving simulators. If driving in a simulator can provide similar physiological responses to on-road driving and more representative physiological responses than those observed using a video-based tasks, this could indicate the validity of simulator driving for testing driving abilities. A number of studies have explored the physiology of driving, measuring variables such as heart rate and skin conductance.

Dijksterhuis, Brookhuis, and De Waard (2011) recorded the physiology of drivers as they participated in a simulated driving study investigating steering and lane keeping. Participants drove in a driving simulator, while the width of the road and the density of oncoming traffic was manipulated. Compared to baseline, driving in the simulator resulted in an increase in heart rate. Meanwhile an increase in the amount of incoming traffic resulted in a decrease in heart rate variability. The authors attributed these changes in heart rate to increased mental workload during the simulated driving.

Skin conductance has also been reported to be sensitive to levels of arousal and workload during driving (Healey and Picard, 2005). Taylor (1964) recorded skin conductance while people drove on the road. The results of the study revealed that skin conductance was negatively correlated with driving speed. He observed that in high-risk situations (when skin conductance levels were high) participants slowed down, whereas in low-risk situations (when skin conductance levels were low) participants drove faster. These were taken to support the idea of risk homeostasis is utilised in driving. Risk homeostasis (Wilde, 1994) is the idea that people attempt to find a balance between their subjective perception of risk and the amount of risk they perceive is acceptable. With regards to driving, this means that drivers will adjust their driving behaviour to balance out any perceived risk. For example, if a driver perceives the level of risk to be lower than what they deem acceptable then they will engage in a behaviour that increases the perceived level of risk (e.g. speeding up). Meanwhile, if a driver perceives the level of risk to be too high they will engage in a behavior that will reduce their level of perceived risk (e.g. slowing down). The fact that perceived risk can result in anxiety allows for measurement via skin conductance.

3.3 Study 2

The aim of Study 2 is to explore if there are any differences in behavioural and psychophysiological responses to hazard videos and simulated driving hazard scenarios, in order to explore the possibility of introducing simulated elements into HP testing, as this may provide a more holistic and representative test of on-road HP ability. Hazard scores will be analysed within specific scoring windows with button responses being collected both within and outside of the hazard windows, while the psychophysiological measures will be collected throughout the duration of the videos and simulations. The purpose of this is to see how the different physical and psychological processes of watching videos and driving in simulated environments affect psychophysiological responses. A further aim is to see if taking one type of HP test (video or simulated) affects performance on subsequent HP test. This second aim is one that was introduced opportunistically (since the basic experiment was relatively short) but is one that is theoretically interesting. Clearly one of the expectations from hazard perception training is that video-based training should transfer into behaviour in real-world driving (e.g. Chapman et al., 2002). The current design thus allows us to additionally explore the degree to which video-based testing improves performance in conceptually similar simulated drivers. The basic design involves people taking a standard video-based HP test or driving analogous situations in a simulator while additionally pressing a response button on the steering wheel in the same way as a response button is used standard HP testing.

It is predicted that psychophysiological responses will be higher for the driving simulations. Additional predictions are that HP scores will be higher for video hazards but that there may be more button responses to the simulated hazards. These predictions are based upon the theory that the simulated HP test will require greater levels of mental and psychophysiological demand and as the participants are in a more immersive and demanding environment they may respond more frequently to potential hazards in the environment. There may also be some evidence of transfer, with the completion of one HP test improving performance on another (e.g. completing the video HP test may improve performance on the simulated HP test or vice-versa). Overall, the simulated hazard scenarios should produce significantly different behavioural and psychophysiological responses than the video hazards.

3.4 Methodology

3.4.1 Participants

A power analysis revealed that for a large effect size (F = 0.4), a total of 32 participants (eight for each of the four participants groups) would be needed for this study.

The participants were 32 drivers who had at least a provisional driving licence and had normal or corrected-to-normal vision. All of the participants were students at the University of Nottingham. 11 of the participants were males. The average age of the participants was 22.6 years (with a range of 13 years). Their annual average mileage was 4245 miles (with a range of 20,000 miles). The average number of hours driven per week was 4 hours (with a range of 15 hours). Participants were paid £5 for their participation, or received credit through the School of Psychology's Research Participation Scheme.

3.4.2 Apparatus and Stimuli

3.4.2.1 Driving Simulator

The experiment took place in NITES2 (the Nottingham Integrated Transport and Environment Simulation facility) driving simulator. This is a fixed base driving simulator with a 180 degrees semi-circular projection screen 2 metres in height and 5 metres in diameter. A rear view was provided by a flat screen television placed behind the participant that could be viewed with a conventional rear view mirror (no side mirrors were present). This simulator provides all standard car controls, with the exception of a handbrake. Hazard responses were made via a pushbutton attached to the steering wheel (for both video and simulated hazards). For standard video-based hazard scenarios, participants remained in the simulator, but the rear screen was wheeled in front of them and videos were shown via a Sony DSR-11 MiniDV digital video recorder. When placed in front of participants the screen subtended approximately 20 degrees of visual angle – similar to what would be the case in a standard Hazard Perception test.

3.4.2.2 Physiological Equipment

The physiological data was recorded using a Biopac MP150. It consisted of a bio potential amplifier, which received input from three separate pre amplifiers (GSR100C, PPG100C and RSP100C). The GSR100C was connected to two skin conductance transducers (TSD203), which had isotonic recording electrode gel 101 and were attached to the participant's fingers. The participants pulse was recorded using a finger mounted optical pulse transducer (TSD200) connecting to the PPG100C pre amplifier.

3.4.2.3 HP Videos

The stimuli consisted of 8 hazard perception videos provided by the UK Driving Standards Agency. These were taken from the licensed training material available with the Hazard perception test and each included one predetermined hazardous event. These were already coded with a hazard onset and offset time, which are used to score performance on the UK hazard perception test. A button press in the first fifth of this time window provided the candidate with a score of 5, while scores in later windows receive lower scores, down to 1 for the last window and a score of 0 for presses that are made before the hazard onset, or after the hazard offset. Hazards videos lasted between 49 and 69 seconds (mean 59.9 seconds). Hazard window lengths lasted between 2 and 12 seconds (mean 6.7 seconds). Descriptions of the 8 hazards used can be found below in Table 3.1 and detailed timings can be found in table 3.2. All of the HP videos contained driving scenes that took place during the day and in dry conditions.

The official introduction from the UK HP test was also used in this study, for the purpose of explaining what the participants were required to do during the video-based HP test. The video contains a video commentary, which defines what a hazard is ("any object that may cause you, as the driver, to slow down or change direction) as well as explaining when the participant is required to do during the test. It also provides an example HP video clip, which is run through twice. The first time any potential driving hazards are circled in yellow, while the score-able hazard is circled in red. The video also provides information on the scoring window for the hazard and their reaction time in responding to the hazard is important, with responses falling outside of the window not receiving a score. The hazard is then repeated, this time without anything being circled, in order to give the participant the opportunity to practise responding to the driving hazard.

Table 3.1. Description and order of the video and simulated hazards used in the experiment, as well as the speed limits for those hazards.

Hazard	Description	Speed limit
1	White van pulls out into the road in front of the car from a left hand side street	60 mph
2	On the motorway, a car comes from the slip road and into the left hand lane in front of the car	70 mph
3	Whilst taking a left hand turn into a suburban road with cars parked on either side of the road, a car enters the street from the right hand side	20 mph
4	When driving behind a motorcyclist, a bike pulls out from the pavement onto the road from the left hand side behind a red land rover, the motorcyclist proceeds to brake suddenly	30 mph
5	A pedestrian is walking on the left pavement beside a parked white van. They then step out into the road from behind the van and cross the road towards another pedestrian	20 mph
6	When following a cyclist who has come from a side street, a parked car then pulls out and forces the cyclist to swerve to avoid it	30 mph
7	A motorbike pulls out into the road heading towards you from a side street on the left. It passes a parked car on the side of the road by moving into your lane	30 mph
8	When following a car in front, a cyclist pulls out into the road in front of the car causing it to brake and slow down.	30 mph

Table 3.2. List of total hazard clip lengths and scoring window lengths for the eight video hazards.

Hazard	Total Clip Length (s)	Hazard Window Length (s)
1	58.25	5
2	69.56	11.44
3	55.48	3.96
4	49.52	2.2
5	57.72	2.3
6	69.28	8.76
7	59.72	7.96
8	59.88	12.04

3.4.2.4 Driving Simulations

Simulated versions of these eight hazards were also created for use in the driving simulator. In each case the behaviour of other vehicles in the simulator was matched to those in the video-based hazard. No attempt was made to precisely match the exact environmental geometry of the videoed hazards; instead we focused on ensuring the visual precursors to the hazard we accurately simulated. Because a simulated drive will always vary from participant to participant depending on their driving behaviour, we could not match exact hazard timings, instead we defined fixed hazard onset and offset locations (in the simulated driving space) that would match those in the videos where the driver was proceeding at the prevailing speed limit. Although rear-view information was available in the simulated condition, no relevant vehicles were present behind the driver in any of the simulated hazards. The simulated hazards were designed to have the same total and scoring window durations as the video hazards when driven through at the speed limit. However, due to differences in the speed at which people drove in driving simulator, there were slight differences in these values for the participants. All of the driving simulations took place in daytime with dry conditions.

3.4.2.5 Simulator Sickness Questionnaire

A standardised simulator sickness questionnaire was used to assess simulator sickness levels in the participants. It can be found in Appendix 3.1.

3.4.3 Notes

In order to provide scoring windows for the HP videos, onsets and offsets of the hazards were based on timings taken from the DSA training clips (which consist of the circled portion of each training clip). The hazard windows were marked on the videos by an audio tone that is sent to the Biopac. Reaction times were taken from the additional button on the steering wheel that is linked to a separate recording channel on the BIOPAC.

For the simulated hazards, each hazard is synched to the BIOPAC by the participant pressing the right indicator until the light on the dash comes on, and then the left indicator to extinguish it. This signal provides a zero time in the simulator data to link to a zero time in the BIOPAC data. Hazard onset and offset are defined by lines in simulated space, which are crossed at the spatial start and end of each hazard. A Matlab script was used to extract the time that the participant crossed each of these lines from the simulator output. This is then compared to the button press data from the BIOPAC to score hazard performance.

3.4.4 Design

The core design was a mixed ANOVA, with the within subjects factors being phase of testing (first vs. second), the 8 different video or simulated driving hazards and the Hazard Window (inside vs. outside). The between subjects factor was the Test Combination (Video-Video, Video-Simulation, Simulation-Video, Simulation-Simulation). Any main effect of phase of testing can be interpreted as showing the degree to which people learn from doing a hazard perception test, and the degree that this interacts with test combination gives a measure of near and far transfer between tests.

The first dependent variable of interest was reaction time after hazard onset for participants to press the response button. This was measured as a pure reaction time, and also a score, with early responses receiving 5, and later responses getting lower values following the algorithm used by the UK Driving Standards Agency. A further dependent variable was the number of button responses (per second) made during each HP video or simulation. The number of button presses inside and outside of the hazard scoring windows was recorded. The other dependent variables were based upon the physiological recordings, skin conductance and heart rate, taken during the tasks. Skin conductance was calculated in terms of the average skin conductance and the number of electrodermal responses that occurred throughout the hazard simulations or videos. The heart rate values recorded were the average heart rate throughout the individual driving simulations or videos, as well as the change in heart rate.

3.4.5 Procedure

The participants first watched the HP introduction video that is present in the UK HP test and provides a detailed explanation of what the participants are required to do during the video hazards. The participants then performed a short drive in the simulator to familiarize themselves with the car controls, experience the simulated driving world and practise responding to a hazard (both through the driving controls and the button on the steering wheel). Once the participants were comfortable with what the tasks involved, the physiological recording devices were put in place. The skin conductance monitors were placed on the index and third fingers of their left hand. The participants then completed two HP tests (either Video-Video, Video-Simulation, Simulation-Video, Simulation-Simulation). Hazards were presented in a single random order, which was the same for all participants and both phases and modalities of testing (the order of the hazards is listed in Table 1).

For the video hazards the participants simply had to sit in the driving simulator, watch the video hazards as they were presented on the TV and respond to any hazards they perceived using the steering-wheel mounted response button. They were told to follow the instructions they had been provided with in the HP introduction video and to press the steering wheel mounted response button when they thought a hazard was present on screen. The eight hazard videos were presented one after the other, with a short gap between each video that consisted of a blank screen. In total, the video hazard condition lasted for approximately 10 minutes, including the short gaps between each of the eight video hazards.

For the simulated hazards, the participants were required to drive through the simulated hazard scenarios using the simulator controls. The participants were instructed to follow the road ahead unless instructed to turn left or right (which was communicated through a pre-recorded message that was set to play at specific locations during the simulated drives). The participants were also instructed to take avoiding action to avoid having a collision with any objects in the road as well as to respond to any hazards using the steering wheel mounted response button. The participants were also informed of the speed limit for each simulated hazard and were instructed to try and drive at the speed limit. After each hazard scenario had been completed a short break was taken during which time the participants completed the simulator sickness questionnaire (Appendix 3.1). If at any point scores of 'Moderate' sickness were reported then the experiment was paused and the participant was provided with some water. The participants were given several minutes after which they were asked if they wanted to continue. If they said that they felt unable to continue then they were removed from testing. If the participants stated that they now felt O.K. then the experiment continued. If at any point a participant reported a score of 'Severe' on any of the measures then they were provided with some water and they were removed from testing. A fan blowing cool air was left on at all times to help reduce incidences of simulation sickness. No participants were removed from testing due to simulator sickness in this study. If the participants drove at the speed limit then the simulated hazard condition lasted for approximately 20 minutes, including the breaks taken between each of the eight simulated hazards to complete the simulator sickness questionnaire.

If the participants crashed at any point during one of the simulated hazards then the scenario was restarted unless the crash occurred very close to the end of the simulation.

3.5 Results

3.5.1 Pre-Analysis

Before the data was fully analysed, a small amount of missing physiological data was replaced by cell means. This consisted of replacing missing data with the weighted average of the participant and condition means. The number of missing values replaced was as follows: 1.2% for button responses, 1.2% for mean skin conductance total.

3.5.2 Button Responses

For hazard and the window by hazard interaction, Mauchly's Test of Sphericity was breached. As a result, Greenhouse-Geisser corrections were used.

The number of button responses per second were analysed using a mixed 4x2x2x8 ANOVA (condition x hazard window x phase x hazard). The results revealed a significant main effect of window, $F(1,28)= 93.458$, $MSE=16.758$, $p<.001$, ${}_{p}\eta^2 =.769$, with significantly more button responses per second being made inside the hazard windows than outside. There was also a significant main effect of Hazard, $F(2.326,65.118)= 16.821$, $MSE=1.328$, $p<.001$, ${}_{p}\eta^2 =.375$, with a significant difference in the number of button responses per second for the 8 hazards. A full breakdown of button responses can be seen in Table 3.2 and Table 3.3. A significant interaction between phase and condition was found, $F(3, 28)= 5.018$, $MSE=0.199$, $p<.05$, ${}_{p}\eta^2 =.350$, which can be seen in Figure 3.1. There was also a complex hazard by condition interaction, $F(21,28)= 1.676$, $MSE=0.044$, $p<.05$, ${}_{p}\eta^2 =.152$, which can be seen in Figure 3.2. Finally, a significant window by hazard interaction was found, $F(2.960,82.883)= 13.017$, $MSE=0.726$, $p<.001$, ${}_{p}\eta^2 =.317$, which can be seen in Figure 3.3 and reveals that while button responses were significantly higher inside of the scoring window for each of the hazards, this difference was not consistent across hazards. A t-test comparing the mean number of total button responses revealed significantly more button presses for the video (mean=6.52, S.D.=1.6) hazards than the simulated (mean=4.37, S.D.=1.46) hazards, $t(62)=5.557$, $p<.001$.

Table 3.3. Average button presses per second for the VV and VS conditions. Values in parentheses are standard deviations.

Condition	VV				VS			
Phase	1		2		1		2	
Window	In	Out	In	Out	In	Out	In	Out
Hazard								
1	0.2 (0.1)	0.07 (0.04)	0.29 (0.18)	0.06 (0.04)	0.17 (0.12)	0.09 (0.08)	0.44 (0.25)	0.06 (0.09)
2	0.22 (0.1)	0.04 (0.03)	0.26 (0.12)	0.03 (0.03)	0.24 (0.16)	0.06 (0.06)	0.19 (0.19)	0.06 (0.15)
3	0.47 (0.16)	0.09 (0.05)	0.47 (0.2)	0.06 (0.06)	0.53 (0.3)	0.08 (0.1)	0.39 (0.36)	0.08 (0.17)
4	0.62 (0.2)	0.13 (0.05)	0.62 (0.33)	0.1 (0.07)	0.79 (0.4)	0.19 (0.22)	0.36 (0.5)	0.06 (0.06)
5	0.34 (0.3)	0.1 (0.05)	0.28 (0.23)	0.1 (0.07)	0.51 (0.9)	0.14 (0.21)	0.3 (0.28)	0.09 (0.1)
6	0.31 (0.13)	0.09 (0.04)	0.38 (0.23)	0.08 (0.06)	0.54 (0.45)	0.07 (0.07)	0.35 (0.35)	0.06 (0.06)
7	0.34 (0.11)	0.07 (0.03)	0.34 (0.13)	0.05 (0.03)	0.28 (0.24)	0.1 (0.1)	0.41 (0.29)	0.07 (0.06)
8	0.16 (0.06)	0.08 (0.04)	0.19 (0.08)	0.08 (0.05)	0.27 (0.24)	0.09 (0.14)	0.12 (0.09)	0.1 (0.1)

Table 3.4. Average button presses per second for the SV and SS conditions. Values in parentheses are standard deviations.

Condition	SV				SS			
Phase	1		2		1		2	
Window	In	Out	In	Out	In	Out	In	Out
Hazard								
1	0.46 (0.33)	0.01 (0.01)	0.27 (0.14)	0.06 (0.07)	0.17 (0.35)	0.05 (0.04)	0.55 (0.68)	0.05 (0.06)
2	0.08 (0.06)	0.01 (0.02)	0.21 (0.13)	0.04 (0.03)	0.12 (0.15)	0.01 (0.02)	0.11 (0.12)	0.01 (0.02)
3	0.34 (0.16)	0.03 (0.04)	0.53 (0.25)	0.07 (0.04)	0.27 (0.38)	0.02 (0.02)	0.25 (0.33)	0.02 (0.02)
4	0.23 (0.2)	0.05 (0.03)	0.55 (0.3)	0.12 (0.09)	0.39 (0.16)	0.05 (0.06)	0.37 (0.32)	0.05 (0.05)
5	0.13 (0.14)	0.04 (0.04)	0.34 (0.21)	0.09 (0.06)	0.37 (0.19)	0.03 (0.04)	0.4 (0.17)	0.04 (0.05)
6	0.18 (0.08)	0.03 (0.02)	0.39 (0.21)	0.06 (0.03)	0.19 (0.13)	0.04 (0.04)	0.2 (0.2)	0.03 (0.03)
7	0.34 (0.16)	0.05 (0.03)	0.33 (0.22)	0.06 (0.05)	0.37 (0.14)	0.06 (0.07)	0.36 (0.33)	0.05 (0.05)
8	0.16 (0.1)	0.06 (0.05)	0.18 (0.12)	0.09 (0.08)	0.08 (0.11)	0.06 (0.06)	0.36 (0.31)	0.05 (0.05)

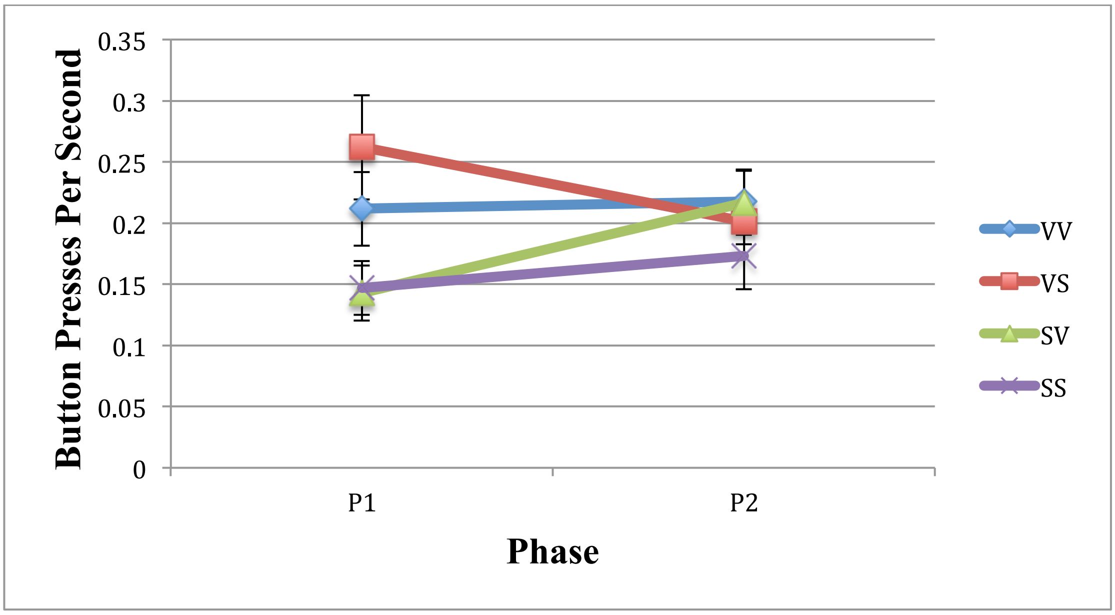


Figure 3.1. Significant interaction between condition and phase for button presses per second. Error bars represent +/- 1 S.E.M.

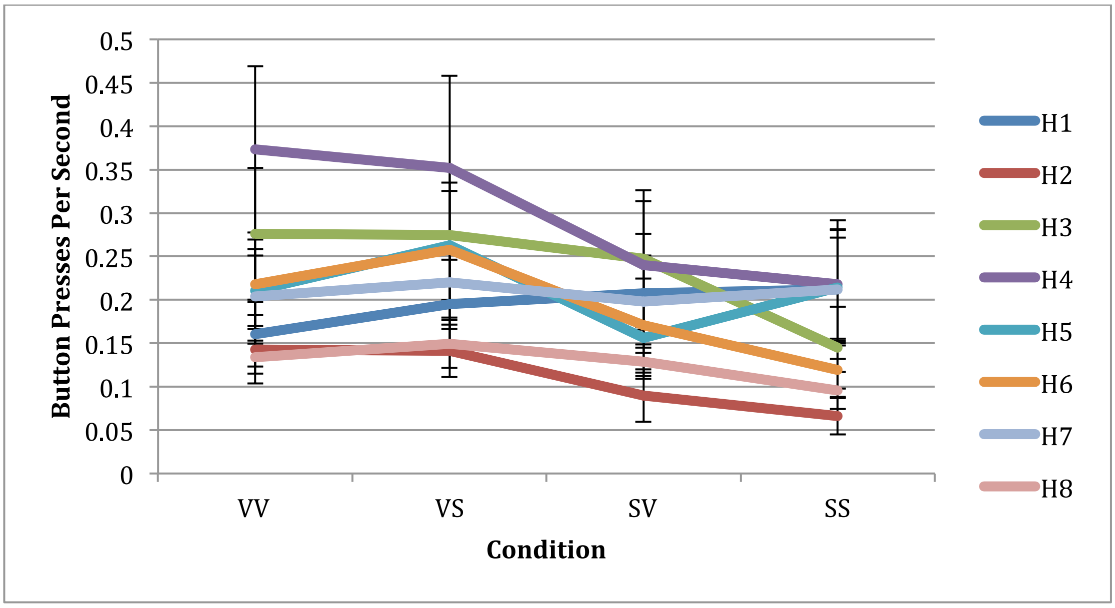


Figure 3.2. Significant interaction between condition and hazard for button presses per second. Error bars represent +/- 1 S.E.M.

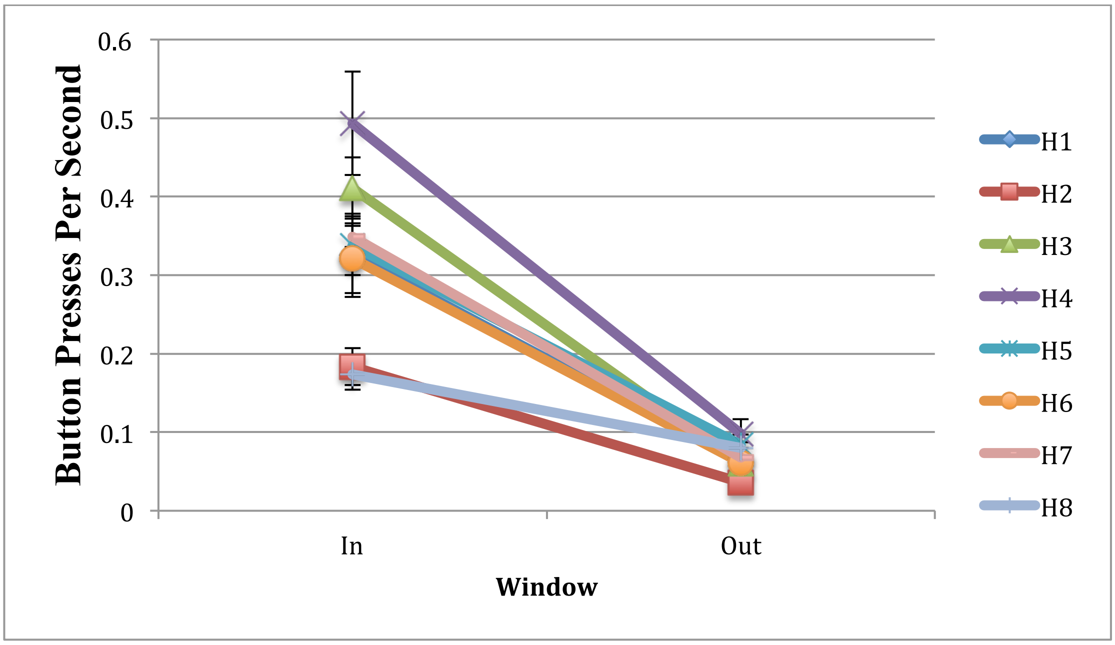


Figure 3.3. Significant interaction between hazard and window for button presses per second. Error bars represent +/- 1 S.E.M.

3.5.3 Hazard Scores

Mauchly's Test of Sphericity was breached for hazard. Therefore Greenhouse-Geisser corrections were used.

Hazard scores were analysed using a 4x2x8 mixed ANOVA (condition x phase x hazard). The results of the ANOVA revealed a significant main effect of Hazard, $F(4.482,125.5)= 20.431$, $MSE=60.347$, $p <.001$, $_p\eta^2 =.422$, with significant differences in the scores for the 8 different hazards. A full breakdown of hazard scores can be seen in Table 3.2. There was also a significant interaction between phase and condition, $F(3,28)= 4.666$, $MSE=12.179$, $p <.05$, $_p\eta^2 =.333$, which can be seen in Figure 3.4. A significant interaction between hazard and condition was also found, $F(21,28)= 5.643$, $MSE=2.984$, $p <.05$, $_p\eta^2 =.242$, which can be seen in Figure 3.5. A t-test comparing mean hazard scores revealed that scores were significantly higher for the video (mean=3.34, S.D.=1) hazards than the simulated (mean=2.46, S.D.=0.88) hazards, $t(62)=3.596$, $p<.01$.

The lack of a main effect of phase indicates that there is no positive transfer of hazard score across the phases. However, the phase by condition interaction indicates that there could be some positive transfer for some of the conditions. In order to see if there was any evidence of positive hazard score transfer, two 2x2x8 (condition x phase x hazard) ANOVAs were carried out. One compared the VV and SS conditions while the other compared the VS and SV conditions. A Bonferroni corrected alpha level of 0.025 was used to account for multiple comparisons.

The results of the ANOVA comparing the VV and SS conditions revealed no significant main effect for phase [$F(1,14)= 0.104$, $MSE=0.141$, $p =.751$, $_p\eta^2 =.007$] and no significant phase by condition interaction [$F(1,14)= 0.940$, $MSE=1.266$, $p =.349$, $_p\eta^2 =.063$], indicating that there was no transfer of scores in these conditions. The ANOVA comparing VS and SV revealed no significant main effect of phase [1,14)= 0.082, $MSE=0.316$, $p =.779$, $_p\eta^2 =.006$] but a significant interaction between phase and condition $F(1,14)= 9.099$, $MSE=35.254$, $p >.01$, $_p\eta^2 =.394$, indicating positive transfer for the SV condition. However, it is important to note that this effect may in fact be due to differences in the baseline performance between the simulated and video hazards.

In order to establish if the ordering of hazard scores was the same for the videos and the simulations (i.e. if the video and simulated hazards are comparable to one another in terms of the responses they elicit), a Spearman's rank-order correlation was carried out using the hazard scores. The results revealed a strong and significant correlation between the video and simulated hazards $r_s(6) = .743$, $p<.05$, demonstrating that the ordering of the hazard scores was very similar for the video and simulated hazards.

Table 3.5. Mean hazard scores for each phase of each condition. Values in parentheses are standard deviations.

Condition	VV		VS		SV		SS	
Phase	1	2	1	2	1	2	1	2
Hazard								
1	1.8 (0.9)	2.2 (1.6)	1.7 (1.7)	2.6 (1.4)	2.6 (1.8)	3.1 (1.5)	0.8 (1.6)	2.1 (2.1)
2	4.1 (1.3)	4.1 (0.8)	3.1 (1.5)	1.7 (1.2)	1.6 (1.1)	3.7 (1.7)	1.6 (1.7)	2 (1.4)
3	4.1 (0.8)	4.1 (0.9)	3.6 (1.1)	2.5 (1.4)	3.3 (1.5)	4.3 (0.5)	1.6 (1.7)	2 (2.1)
4	3.7 (0.7)	3.1 (1.3)	3 (1.5)	3.1 (2)	2.1 (2.1)	3 (1.7)	3.1 (1.7)	3.8 (1.7)
5	2.3 (2.3)	2.1 (2.2)	1.5 (2)	1.6 (1.5)	2 (2.2)	1.7 (1.7)	3 (2)	2.5 (1.7)
6	4.8 (0.3)	4 (1.9)	5 (0)	3.3 (2.2)	4.7 (0.7)	5 (0)	3.7 (1.4)	3.2 (1.8)
7	4.3 (1)	4.5 (1.1)	3.2 (2)	2.1 (2.1)	3.1 (1.7)	5 (0)	3.1 (1.8)	2.7 (2.2)
8	2.1 (0.6)	2.6 (0.7)	2.5 (1.4)	1.25 (1.8)	2.7 (1.4)	2.8 (1.2)	1.2 (1.9)	1.3 (1.9)

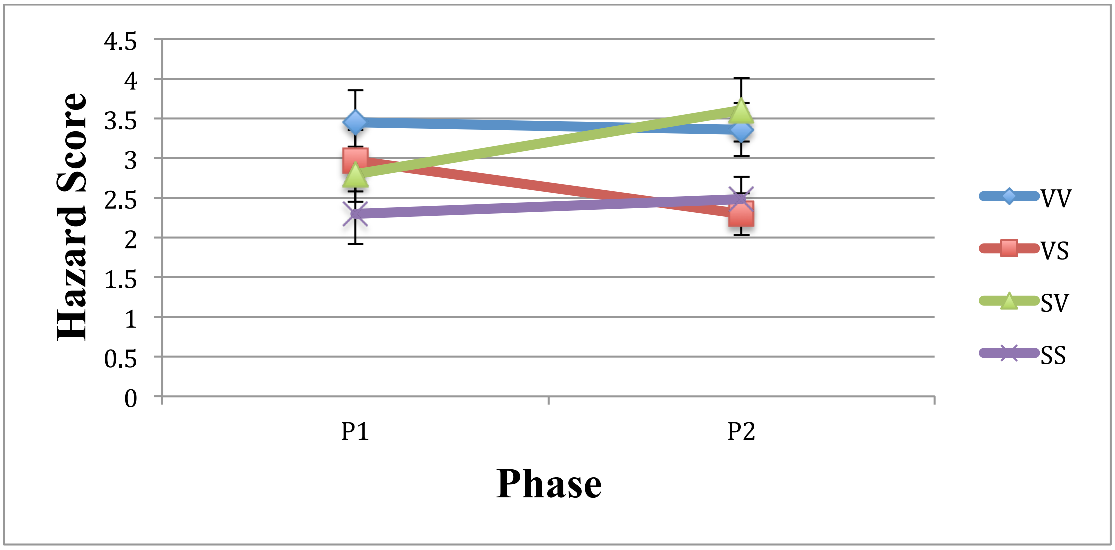


Figure 3.4. Significant interaction between condition and phase for hazard scores. Error bars represent +/- 1 S.E.M.

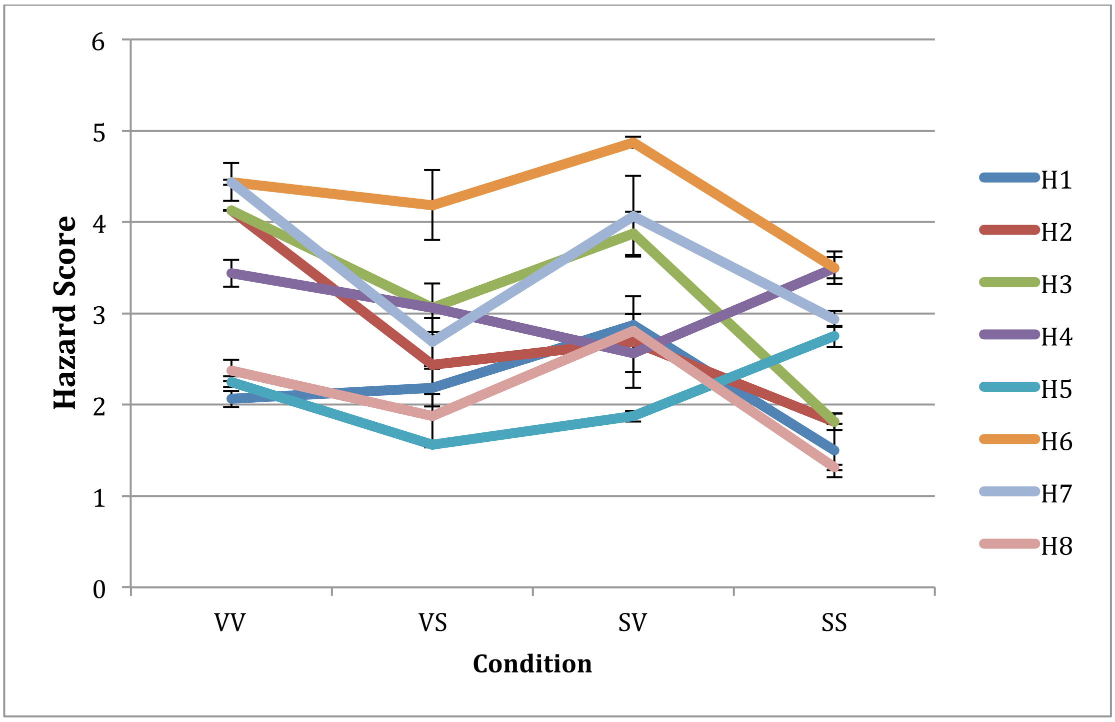


Figure 3.5. Significant interaction between condition and hazard for hazard scores. Error bars represent +/- 1 S.E.M.

3.5.4 Skin Conductance

Mauchly's Test of Sphericity was breached for hazard and the phase by hazard interaction. Therefore Greenhouse-Geisser corrections were used.

Mean skin conductance total was analysed using a 4x2x8 (condition x phase x hazard) mixed ANOVA. A significant main effect of hazard was found, $F(2.721,76.186)= 11.468$, $MSE=25.673$, $p <.001$, ${}_{p}\eta^2 =.291$, with significant differences in the mean skin conductance for the 8 hazards. This can be seen in Table 3.6. There was also a significant interaction between phase and condition, $F(3,28)= 16.44$, $MSE=139.994$, $p <.001$, ${}_{p}\eta^2 =.638$, which can be seen in Figure 3.9. A significant interaction between hazard and condition was also found, $F(21,28)= 1.627$, $MSE=1.415$, $p <.05$, ${}_{p}\eta^2 =.148$, which can be seen in Figure 3.10. A significant interaction between hazard and phase was found, $F(2.116,58.258)= 3.303$, $MSE=11.451$, $p <.05$, ${}_{p}\eta^2 =.106$, which can bee seen in Figure 3.11. A t-test revealed that mean skin conductance was higher for the simulated hazard scenarios (mean=9.59, S.D.=0.95) than the hazard videos (mean=7, S.D.=0.87), $t(62)=11.27$, $p<.001$.

Table 3.6. Mean skin conductance for each hazard in each phase of each condition. Values in parentheses are standard deviations.

Condition	VV		VS		SV		SS	
Phase	1	2	1	2	1	2	1	2
Hazard								
1	7.2 (3.5)	6.8 (4.5)	7.9 (3.4)	10.5 (3.9)	8.6 (2.2)	8.9 (2.9)	10.8 (2.9)	9.9 (1.9)
2	6.6 (3.3)	6.5 (4)	7.7 (3.3)	10.5 (3.9)	8.5 (2.1)	8.5 (2.6)	11.1 (3.1)	9.8 (2)
3	6.2 (3.2)	7 (4.8)	7.2 (3)	11.1 (4.3)	9 (2.3)	8.3 (2.8)	9.8 (3.7)	8.9 (2.2)
4	6.2 (3.5)	6.2 (4)	7.3 (3.7)	10.6 (3.7)	9.1 (2.4)	8.2 (2.8)	10.6 (2.9)	8.9 (1.9)
5	6.1 (3.7)	6.2 (4.1)	7.1 (3.3)	10.7 (3.7)	9.6 (2.9)	8.2 (3)	9.7 (2.7)	8.6 (1.8)
6	6.2 (4)	6.1 (4)	6.7 (3.3)	9.9 (3.2)	9.3 (3)	7.9 (2.9)	9.6 (1.9)	7.8 (2)
7	6.1 (4.1)	5.7 (3.8)	6.8 (3.6)	9.6 (3.7)	9.2 (3.1)	7.4 (2.1)	10.2 (2.4)	7.6 (1.9)
8	6.2 (3.9)	5.5 (3.6)	6.8 (3.5)	9.6 (3.5)	9.4 (3.2)	7.3 (2)	9.9 (2.1)	7.5 (2.2)

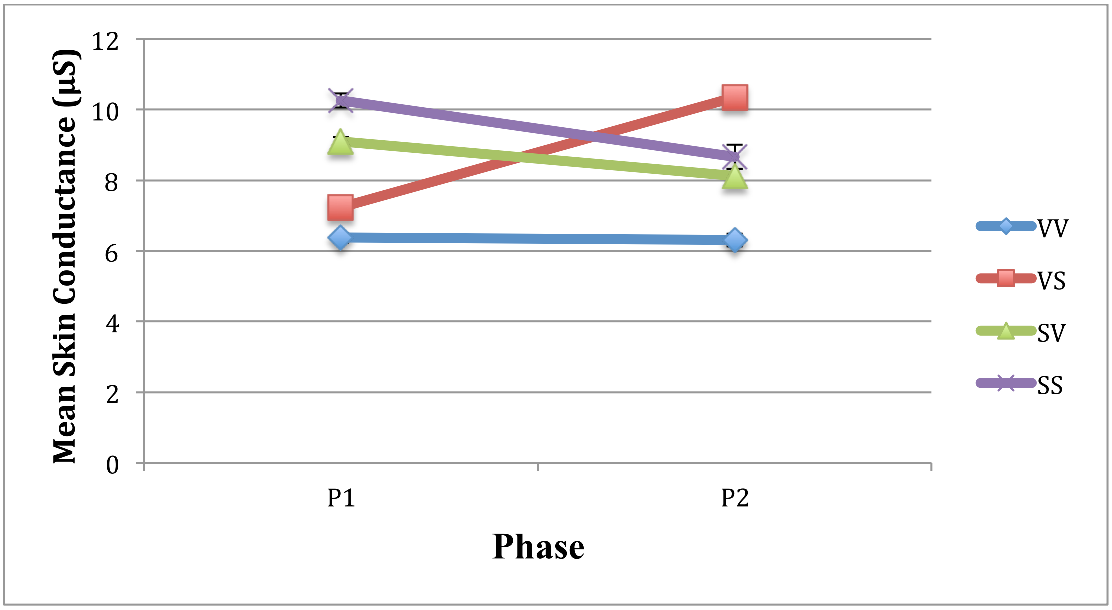


Figure 3.6. Significant interaction between phase and condition for mean skin conductance. Error bars represent +/- 1 S.E.M.

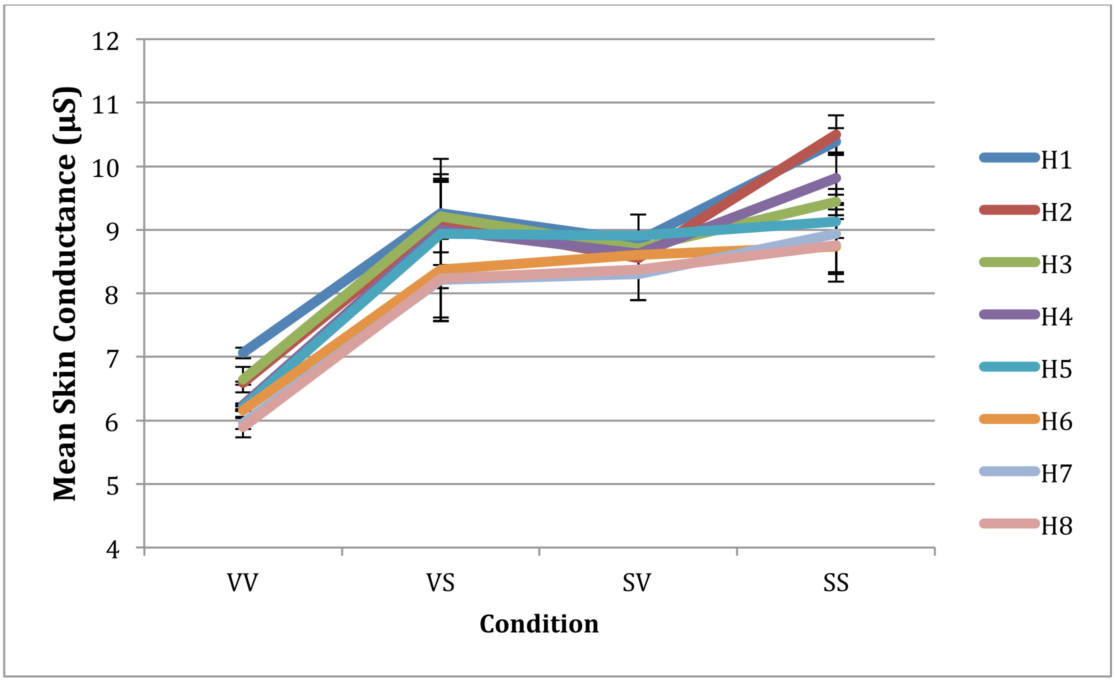


Figure 3.7. Significant interaction between condition and hazard for mean skin conductance. Error bars represent +/- 1 S.E.M.

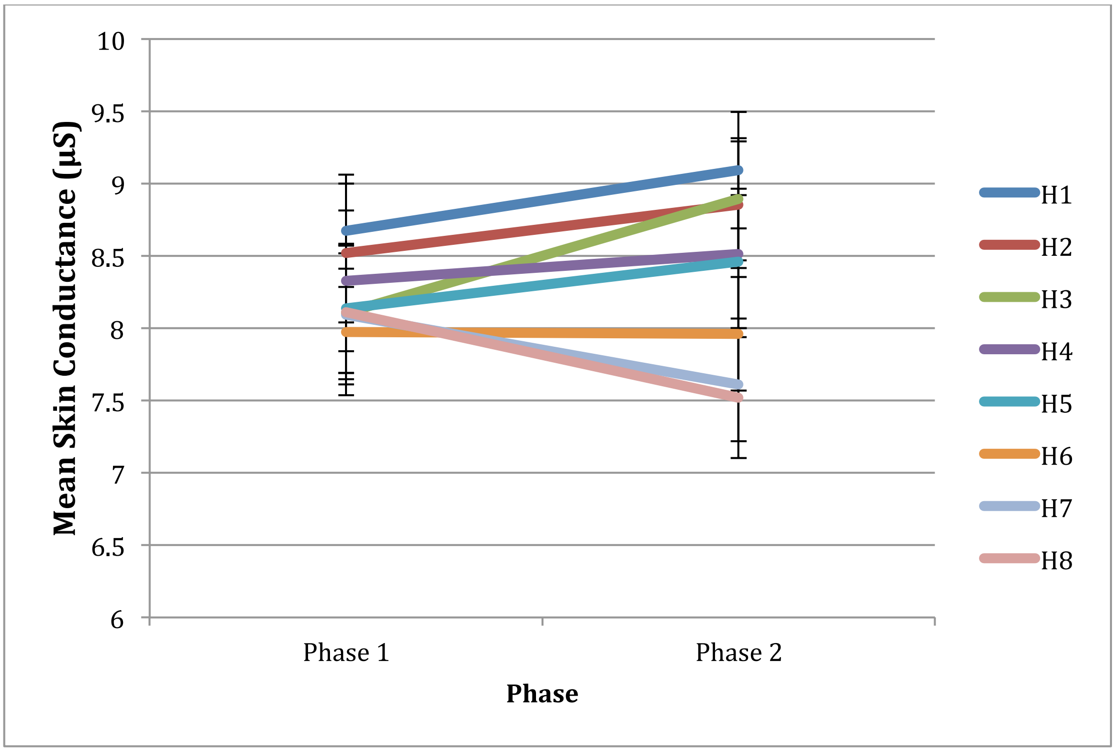


Figure 3.8. Significant interaction between phase and hazard for mean skin conductance. Error bars represent +/- 1 S.E.M.

3.4.6 Heart Rate

The mean heart rate was analysed using a 4x2x8 (condition x phase x hazard) mixed ANOVA. Analysis of the heart rate data failed to reveal any significant main effects or significant interactions. A summary of the mean heart rate data can be found in Table 3.8.

Table 3.7. Mean heart rate for each hazard in each phase of each condition. Values in parentheses are standard deviations.

Condition	VV		VS		SV		SS	
Phase	1	2	1	2	1	2	1	2
Hazard								
1	88 (18)	87 (19)	73 (21)	80 (13)	81 (17)	76 (10)	81 (17)	78 (18)
2	85 (18)	86 (17)	68 (22)	80 (16)	83 (15)	77 (13)	78 (18)	77 (19)
3	82 (16)	84 (18)	67 (24)	78 (9)	80 (16)	75 (14)	80 (11)	84 (10)
4	83 (17)	81 (13)	73 (18)	75 (14)	82 (17)	74 (9)	79 (18)	75 (12)
5	84 (15)	84 (19)	77 (16)	79 (12)	79 (18)	76 (12)	81 (10)	79 (17)
6	83 (15)	81 (17)	74 (15)	75 (12)	78 (14)	76 (13)	80 (18)	73 (16)
7	81 (12)	80 (17)	76 (15)	75 (14)	78 (14)	75 (9)	82 (18)	81 (11)
8	84 (16)	82 (18)	66 (22)	77 (16)	79 (15)	75 (8)	81 (15)	76 (17)

3.5 Discussion

The aim of Study 2 was to investigate whether simulated hazard scenarios produced significantly different behavioural and psychophysiological responses than hazard videos, in order to explore the possibility of introducing simulated elements into HP testing. The general findings of the experiment are that: participants scored higher and responded more when watching video hazards, while skin conductance and the number of button responses was higher for the simulated hazard scenarios. This demonstrates that the simulated hazard scenarios were successful in producing significantly different behavioural and physiological responses than the hazard videos. The hazard scores and skin conductance results are in line with the predictions of this study but the number of responses and heart rate results were not. There was also potentially some evidence for a positive transfer of HP scores in one of the testing conditions.

The finding that there was significantly more button response inside of the hazard windows is in line with predictions. This demonstrates that the hazards, contained within the scoring windows, functioned as they were designed to do by eliciting more responses from the participants than any of the potential hazards that occurred outside of the scoring windows. The fact that there were significantly more button responses for the video hazards than the simulated hazards is not in line with predictions. It was predicted that the increased realism and interactivity of the simulated HP scenarios would increase the level of hazard experienced, but his was not found. A possible explanation for this is that the significantly higher task workload for the simulated hazards may have interfered with their ability to respond to hazards with the button response while simultaneously taking action with the driving controls to avoid a collision. This is supported by anecdotal evidence from a number of the participants who stated that they had been so focused on the driving controls that they had not had the opportunity to respond as frequently as they did when watching the video hazards. As a result of this, the lower levels of button responding for the simulated hazards may be an indication of the task having been more difficult than that of the video hazards. A further possible explanation is that the simulated hazards were more difficult to identify than their video counterparts, which will have affected the responses to them.

Hazard scores were found to be significantly higher for video hazards than simulated hazards. One explanation for this finding is that the hazard scores were impacted by the significant differences in the number of button responses between the video and simulated hazards. The fact that the participants responded fewer times to the simulated hazards is likely to have impacted on the hazard scores for those simulated hazards. As mentioned above, the fewer button responses made may have been an indication of increased overall task difficulty, which could have impacted scoring performance for the simulated hazards as well. An alternative explanation for this finding could be that simulated hazards were more difficult to identify than the video hazards, which may have also been one of the reasons that the button responses were lower for the simulated hazards. Overall the causes for the lower hazard scores for the simulated hazards may be the same as the reasons for the lower number of button responses. The finding that the ordering of the video and simulated hazards was similar demonstrates that the video and simulated hazards were comparable in nature. This, combined with the findings if increased button responses in the hazard windows, suggests that the simulated hazards were effective and operated as intended.

Skin conductance was found to be significantly higher during the simulated hazard scenarios than during the hazard videos. The increased levels of skin conductance demonstrate that the simulated hazard scenarios produced greater levels of arousal than the hazard videos. This is in line with predictions as the simulated hazard scenarios involve greater levels of physical and psychological workload than the hazard videos. This is largely due to the fact that the simulated hazard scenarios require the participants to physically interact through the use of the simulator driving controls, whereas the hazard videos simply require the participants to watch videos. As research has previously shown, increased skin conductance measured during driving has been associated with an increased perception of risk (Taylor, 1969). This indicates that the increased levels of skin conductance during simulated driving may in part be due to an increased perception of risk on behalf of the participants. The fact that skin conductance was increased but button pressing wasn't in simulated hazard scenarios supports the interpretation of reduced button presses and HP scores in simulated condition as a dual task decrement in performance rather than a failure to identify hazards in simulated conditions.

The failure to find any significant differences in heart rate between the simulated hazard scenarios and hazard videos is surprising and is not in line with the predictions of the study or previous driving research, which has demonstrated that driving results in an increased heart rate, both for real driving (de Waard and Brookhuis, 1991) and simulated driving (Dijksterhuis, Brookhuis, and De Waard, 2011). However, it is important to mention that these previous studies had very different designs and research aims compared to this study, as they studied on-road driving. Given that no studies have yet recorded heart rate while participants complete simulated hazard scenarios and hazard videos, there are no studies that these results can be directly compared with. It is possible that the process of driving in a simulator and responding to hazards simply doesn't evoke an increased heart rate compared to watching and responding to hazard videos. One possible explanation for this is that any increased heart rate may only be associated with the individual scoring hazards in each simulation, which may not be enough to result in significant differences in heart rate recorded across the entire simulation. Ultimately these results are perhaps surprising given that the skin conductance results indicate that the simulated hazard scenarios resulted in increased levels of arousal compared to the hazard videos.

When looking at the hazard scores, the lack of a main effect for phase indicates that there is little evidence of positive transfer across the testing phases (i.e. the experience of competing one HP test having a positive influence on scores in the second HP test). However, the phase by condition interaction indicated that there might have been some positive transfer of scores across phases. After carrying out further tests the results indicated that completing a simulated HP test aided performance on a subsequent video HP test. There was no evidence of positive transfer in any of the other conditions. This result could indicate that undertaking some simulator based HP training could help train learner drivers train and improve their HP abilities. It is however possible that these differences in hazard score may be due to differences in baseline performance between the simulated and video hazards, with lower levels of performance for the simulated hazards than the video hazards.

It is possible that the differences in the behavioural responses to the filmed and simulated hazards are related to differences in the content of the hazard clips. Although as much effort as possible was made to make match the content of the simulated clips to that of the filmed clips, there were inevitably some differences. One of the main differences was the environment. It was simply not possible to build an exact environmental replica of those used in the filmed clips. Instead, the most suitable areas of the existing simulated environment were used, and these were ultimately different from those used in the filmed clips. Another major difference between the simulated and filmed hazards relates to the other vehicles and pedestrians in the driving environment. The simulated hazards were created so that the presence and behaviour of other vehicles and pedestrians in the road environment matched those in the filmed hazards as closely as possible. However, there were again limits as to what was possible simulate using the AI for other vehicles and pedestrians. As a result there were differences in the behaviour of the other vehicles and pedestrians in the simulated environment.

Chapman and Crundall (2011) compared the responses of learner drivers to video hazard perception clips with computer-animated versions of those clips. Directly comparing the button responses between the standard video and animated versions revealed that although the overall pattern of responses was similar, there were some significant differences. Differences in the response rates, peak response rates and the timings at which the responses occurred were found. The authors found that small differences in the content of the video and animated hazard clips were responsible for these differences. The differences included different numbers of pedestrians and parked cars. For example, an extra pedestrian in one of the animated clips may have resulted in the learner drivers attending a different part of the visual scene, altering their responses compared to the standard video version. Differences that occurred near and far from the scoring hazard both affected responses. Although this experiment involved simply watching the hazards as opposed to driving through a simulated environment, it is useful to note that content differences in hazard clips can affect the way people respond to them. It is however important to note that the similar ordering of the video and simulated hazards suggests that any effect caused by differences in the content of the hazards may be relatively small.

The differences in the behavioural and physiological responses to the video and simulated hazards may have important implications for HP testing. A potential criticism of video-based HP tests is that they do not provide a fully representative test of HP abilities as the testing format is significantly detached from the process of detecting hazards while driving on the road. In this study, simulated hazard scenarios were used as a method of better approximating on-road driving. The fact that the simulated hazard scenarios resulted in significantly different behavioural and psychophysiological responses indicates that video-based HP tests may not be evoking the responses associated with detecting hazards in real-world driving. These findings may therefore suggest that video-based HP tests could be improved by implementing simulated driving elements into the tests in order to provide a more realistic test of HP abilities.

One significant limitation of the current research is that by requiring both a button response and car control in the simulated hazard conditions we created a dual task situation that is not required in traditional HP testing, or in actual driving. When most researchers look at hazard responses in a simulator this problem is avoided by using car control actions (braking, steering etc.) as the measure of hazard response. The drawback with this approach is that the timing and scoring of such measures is completely different to that of button pressing in a traditional HP test and means that direct comparisons between environments in terms of HP scores are not possible. It would of course be interesting to have participants drive these simulated HP scenarios without button pressing. It is our expectation that it would be very difficult to create comparable scores based on purely driving measures, but it would nonetheless be interesting to explore the degree to which driving behaviour itself was changed by the secondary button-pressing task.

A further possible explanation for some of the differences relates to the differences in the testing environment itself. Whereas the video hazards were presented on a TV screen measuring, the simulated hazards were presented on a 180 degrees field of view projection screen. This large difference in viewing may have impacted on the results of the study. Clearly this did not allow drivers to detect more hazards in the larger visual environment but the extra degree of visual immersion may have been partly responsible for increased skin conductance.

In conclusion, simulated hazard scenarios produced significantly different behavioural and psychophysiological responses than the hazard videos. The simulated hazard scenarios resulted in some greater psychophysiological responses, as well as reduced behavioural responses and performance on the simulated hazards. There was also potentially some evidence of transfer of HP scores, with simulated hazard aiding performance on subsequent video hazards. The differences in the responses between the simulated and video hazards may be an indication that altering video-based HP tests to include elements that more closely mimic on-road driving may improve and provide a more realistic test of HP abilities; however, they also demonstrate the complexity inherent in using simulators to assess HP. The results of this may also be an indication for the care that is needed when designing simulated hazards, as well as the need for more comparable visual environments when comparing hazard videos to simulated hazards.

Chapter 4 - Introduction to the Prefrontal Cortex

4.1 General Introduction

As was briefly mentioned in Chapter 1, the other major aim of this thesis is to explore what, if any, role the DLPFC plays in driving. This chapter will provide some background information on the physiology of the DLPFC (as well as other prefrontal areas), the cognitive functions associated with these areas and the possible role frontal areas play in driving. This chapter will also introduce the functional neuroimaging method, functional near-infrared spectroscopy (fNIR), which will be used to explore the possible role of the frontal lobes in driving. This will include explaining the technical aspects, procedures, applications and research using fNIR.

4.2. The Prefrontal Cortex

4.2.1 The Physiology of the Frontal Lobes

The DLPFC is located within the frontal lobe. The frontal lobe is one of the four major lobes that make up the cerebral cortex in the human brain. It is located at the front of the brain, separated from the parietal lobe by the central sulcus and the temporal lobe by the lateral sulcus. The frontal lobe contains the following distinct areas:

- Primary Motor Cortex – Brodmann Area 4
- Premotor Cortex (including the Supplementary Motor Area, SMA) – Brodmann Area 6
- Frontal Eye Fields - Brodmann Area 8
- Dorsolateral Prefrontal Cortex – Brodmann Area 9, 46
- Anterior Prefrontal Cortex – Brodmann Area 10
- Orbitofrontal Cortex – Brodmann Area 11, 12
- Insular Cortex – Brodmann Area 13
- Ventral Anterior Cingulate Cortex - Brodmann Area 24

- Ventromedial Prefrontal Cortex (Subgenual Cortex) – Brodmann Area 25
- Dorsal Anterior Cingulate Cortex – Brodmann Area 32
- Anterior Cingulate Cortex – Brodmann Area 33
- Broca's Area – Brodmann Area 44, 45
- Inferior Prefrontal Gyrus – Brodmann Area 47

It has been proposed that the frontal lobes have three major divisions: motor, premotor and prefrontal (Miller and Cummins, 2007). These divisions correspond to the primary motor cortex, premotor cortex (including the supplementary motor area), and prefrontal cortex (PFC). The primary motor cortex is located at the posterior of the frontal lobes and is heavily involved in the planning and execution of movements. The premotor cortex is located just anterior to the primary motor cortex and is thought to be involved in the planning and guidance of movements. It forms BA 6 with the SMA, which contributes to movement control. The PFC is a large area of the frontal lobes consisting of the anterior parts of the frontal lobes. It is involved in a wide range of complex functions including decision-making, problem solving, as well as moderating social behaviour and personality.

The frontal lobes are one of the last areas of the brain to develop, not reaching full maturation until roughly 25 years of age (Sowell, Thompson, Holmes et al., 1999). The frontal lobe contains more dopamine-sensitive neurons than anywhere else in the cerebral cortex, with the mesocortical pathway (a major dopamine pathway) passing through it. The dopamine system is associated with important functions such as reward, attention, planning and motivation. The frontal lobe also contains a large number of connections to other parts of the brain including the other lobes as well subcortical structures.

This chapter will focus on the PFC areas of the frontal lobes over the primary motor and premotor cortex. This is because this thesis is interested in the cognitive aspects of driving rather than the movement and movement planning aspects of driving.

4.2.2 The Functions of the Prefrontal Cortex

The PFC is often divided into three major regions: orbital, medial and lateral. The PFC constitutes nearly one third of the neocortex. The PFC is connected to the brainstem, the thalamus and the limbic system through subcortical networks. The PFC is also connected to the other association cortices, while the lateral, medial and orbital regions are connected with each other. These connections allow the PFC to combine information from a wide range of sources in order to help guide behaviour.

Studies investigating the effects of damage to the PFC have provided a large amount of information regarding their functions. The most famous case of damage to the PFC is that of Phineas Gage, an American railroad constructor who suffered significant damage to his PFC after a railroad spike was driven through his head. This particular case provided the first indication of the important role that the PFC plays in regulating behaviour and personality. Despite surviving the accident, Gage underwent a number of significant changes to his personality and behaviour after the accident. Much of the observations of Gage after his accident came from the physician John Martyn Harlow who published reports in 1848 and 1868. Harlow noted that the changes in Gage's personality and behaviour were so great that people who knew him stated that he seemed to be a different person. Harlow stated, "The equilibrium or balance, so to speak, between his intellectual faculties and animal propensities, seems to have been destroyed. He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operations, which are no sooner arranged than they are abandoned in turn for others appearing more feasible" Harlow (1868, p. 13). However, the significant changes to Gage's personality and behaviour appeared to be temporary, with an account of him in later life stating that he was in good health and suffered from no impairment of his mental faculties (Hamilton, 1860).

A number of studies have used modern day neuroimaging techniques to determine the exact areas of Gage's brain that were destroyed as a result of the accident. Damasio et al. (1994) took exact measurements and created an X-ray of

Gage's skull. From this they created a digital reconstruction of the skull as well as entry and exit points for railroad rod. From this they calculated that the areas of the brain that would have been damaged included both the left and right PFC. In both the left and right hemispheres the orbital frontal cortex suffered damage. The lesion also affected polar and anterior mesial frontal cortices as well as the most anterior section of the anterior cingulate cortex. Overall there was greater damage to the white matter core in the left hemisphere. There was no damage located outside of the PFC.

The behaviours displayed by Gage fit with findings from other studies on lesions to the orbital PFC (Fuster, 1997). People with orbital PFC lesion have been found to be disinhibited and impulsive, which leads to dangerous and risk-taking behaviour. This is due to the fact that the orbital PFC exerts inhibitory control via its connections to the basal ganglia, hypothalamus as well as other areas of the neocortex (Fuster et al. 2000). They also have difficulties sticking to social and moral norms (e.g. by swearing uncontrollably). Those with orbital PFC lesions are also characterised by disorders of attention, being unable to withstand distraction. Lesions of the medial PFC (which includes the anterior regions of the cingulate cortex which was damaged in the case of Phineas Gage) have shown that it is involved in emotion, attention and motility, with afflicted individuals suffering from a difficulty in initiating movements and speech (Devinsky, Morrell and Vogt, 1995). Damage to the cingulate cortex itself results in apathy, disinterest and an inability to concentrate and sustain attention (Devinsky, Morrell and Vogt, 1995). Lesions to the lateral PFC are characterised by the inability to formulate and to carry out plans, as well as the inability to plan and carry out sequences of actions (Luria, 1966). A consistent symptom of lateral PFC damage is planning deficits that affect the representation and construction of sequences of language, both written and spoken (Luria, 1970). This manifests itself in two ways: a difficulty in consciously representing sequences of speech and behaviour, as well as a difficulty in initiating them and executing them. This effect has been termed as the 'dysexecutive syndrome' (Baddeley, 1986). This syndrome is usually accompanied by a severe disorder of attention.

These findings show that the PFC is involved in vital functions such as: inhibition, behavioural regulation, attention, emotion, as well as the planning and execution of functions such as language. Another one of the major functions of the PFC is to provide executive functions. The frontal lobes are often referred to as the 'executive' or 'CEO' of the brain. This is primarily because the PFC is responsible for the executive functions of the brain. Executive function refers to a wide range of cognitive processes to facilitate the initiation, planning, regulation, sequencing and achievement of complex goal-oriented behavior and thought (Shallice, 1988; Stuss and Knight 2002). Norman and Shallice (2000) have proposed that the purpose of executive functions is to deal with novel situations that cannot be dealt with using automatic schemata through simple pattern matching. Executive functions include (Weinberger, Elvevag and Giedd 2005):

- Problem solving and decision making
- Working memory (storage, manipulation and utilisation of complex information)
- Cognitive flexibility (set shifting)
- Planning and strategizing
- Judging risk
- Insight and reasoning
- Allocating attention
- Impulse control and inhibition
- Regulation of behaviour and emotion
- Abstract thinking
- Organisational skills
- Reasoning

A meta analysis of neuroimaging and lesion studies has revealed that the dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) are heavily associated with executive functions (Alvarez and Emory, 2006). Research has shown that the DLPFC is involved in verbal fluency, inhibition, planning, working memory, organisational skills, abstract thinking, reasoning and set shifting (Alvarez and Emory, 2006). The ACC is associated with response inhibition, emotional drive and motivated behaviours, as well as decision- making (Lezak, Howieson and Loring, 2004). The OFC is responsible for impulse control, maintaining set, producing socially appropriate behaviours as well as monitoring behaviour.

This is supported by theories regarding subcortical networks in the frontal lobes. A number of researchers have suggested that there are three major subcortical networks in the frontal lobes that are heavily involved in the cognitive, emotional and motivational processes associated with the frontal lobes: dorsolateral, ventromedial and orbitofrontal (Sbordone, 2000; Stuss and Benson, 1984). The DLPFC projects primarily to the dorsolateral head of the caudate nucleus and has been linked to executive functions, as explained previously. The ventromedial network projects from the anterior cingulate to the nucleus-accumbens. It is thought to be involved in motivation, with lesions to the region resulting in apathy, psychomotor problems and decreased social interaction (Sbordone, 2000). The orbitofrontal network projects from orbitofrontal cortex to the caudate nucleus. This area is thought to be involved in mediating socially appropriate behaviour as lesions to this area result in antisocial behaviour, a lack if inhibition and impulsivity (Cummings, 1995).

The frontal lobes also have connections to other cortical, subcortical and brain stem areas. Alvarez and Emory (2006) stress that executive functions are the result of not just frontal areas, but are a product of these connections between the frontal lobes and cortical and subcortical areas, as well as the brain stem. They argue that the higher cognitive functions that constitute the executive functions are often the result of lower level cognition and behaviours. As a result executive functions need to be thought of with regards to the diverse system of brain areas that have connections to the frontal lobes.

4.2.3 – The Prefrontal Cortex and Driving

By looking at the functions that the frontal lobes provide, it is clear that these functions may have applications in the complex task of driving. There has also been research directly investigating what role the frontal lobe plays in driving. This research has involved a wide range of methods including imaging studies, brain stimulation techniques, as well as studies of the elderly and those with neurological disorders or brain injuries.

4.2.3.1 Imaging Studies

Calhoun et al. (2002) investigated the time-course of neural activity during driving using an fMRI independent component analysis method they developed. Participants took part in two experimental tasks; an active simulated driving task and a task in which they watched a simulated driving scene. In the active simulated driving task the participants used a modified game controller to control a simulated car. They were tasked with staying in one lane unless they needed to overtake or avoid collisions, while keeping their speed within a specified range. The other task involved watching a replay of a previously recorded driving task. They examined changes in neural activity associated with differences in speed as well as starting and stopping driving. The analysis method allowed for the identification of different areas of the brain that displayed specific patterns of activation during driving and watching.

Two of the areas that displayed distinct patterns of activation were orbitofrontal and frontoparietal areas. Orbitofrontal areas (including the anterior cingulate) as well as frontoparietal areas displayed a decrease in activity during the driving task compared to the watching task. Whereas activity in frontoparietal areas (including the posterior cingulate) decreased rapidly during the driving tasks and the rebounded quickly rebound during the watching task, activity in the orbitofrontal decreased and recovered more slowly. The results also revealed speed related changes in orbitofrontal areas, with neural activity in these areas reducing at a faster rate when vehicle speeds were higher. For the orbitofrontal areas, the decrease in activity as a result of faster speeds was more rapid than in the frontoparietal areas.

Calhoun et al. (2002) stated that the orbitofrontal areas they recorded activity from are associated with cognitive functions such as error monitoring, risk assessment and inhibition. The frontoparietal areas meanwhile are associated with cognitive functions such as vigilance, spatial attention and monitoring. They argued that the reductions in neural activity associated with driving were due to reductions in the cognitive processes associated with those particular areas (e.g. reductions in error monitoring result in decreases in orbitofrontal areas). Meanwhile the faster reductions in neural activity associated with increases in driving speed were potentially the result of greater changes in these cognitive components, leading to a faster rate of decrease in neural activity. For example, during fast driving task the vigilance component underwent a greater change, leading to a greater reduction in the error-monitoring and disinhibition components. The results of this study are supported by research from Peres et al. (2000) and Horikawa et al. (2005) who also reported that activity in frontal areas was affected by speed.

Uchimaya et al. (2003) also reported a negative correlation between anterior cingulate activity and task performance during a simulated following task, with reduced anterior cingulate activity associated with driving at a safe distance and increased activity associated with driving at an unsafe distance. The authors attributed this to the role of the anterior cingulate in error detection and response selection.

Spiers and Maguire (2007) investigated the neural substrates of driving with fMRI. 20 London taxi drivers (selected in order to control for navigation experience of London) participated in a simulated driving task. The participants were tasked with driving around accurately simulated areas of central London in a black taxicab. The participants were given seven routes to drive. For each route they were tasked with delivering a passenger to a destination. They controlled the simulated black taxicab through the use of a game controller. The driving simulation required the participants to obey the normal traffic regulations of the road (e.g. stopping at red lights, correctly using one-way streets, allowing pedestrians to cross the road in certain places). The eye movements of the participants were also recorded. The simulated driving environment was shown on a screen that could be viewed by the participants as they were lying in the MRI scanner.

After the scanning session was completed the participants were presented with a video of their simulated driver and were required to provide a verbal report on what they were thinking during each part of their simulated drive. The purpose of this was to provide the timings of the thoughts and mental processes of the participants during the different parts of their drive to help interpret the fMRI data. The researchers assigned the statements from the participants to a number of categories: prepared actions (e.g. turning left), unprepared actions (e.g. avoiding collisions) and planning (monitoring traffic). The eye-tracking data was used to aid the identification of onsets and durations where the participants had reported looking at moving vehicles, pedestrians as well as any fixed features in the environment.

Activity in the supplementary motor area was associated with a number of driving events including, performing planned actions (including starting moving, reversing, turning corners and stopping), planning actions for future events and responding to hazards (e.g. swerving to avoid a collision). The authors also reported that activity the right lateral and medial PFC increased when the participants had to consider the traffic regulations. The authors argued that the activation of the lateral PFC was due to its role in rule retrieval and maintenance. They went on to suggest that the activation of the medial PFC may have been due to the region being responsible for altering responses in order to adapt to challenges in the environment.

Jancke, Brunner and Esslen (2008) investigated the effect that driving speed had on neural activity in the PFC. Participants drove in a driving simulator while PFC activity was monitored using EEG. The participants were given three driving tasks, to drive excessively fast, to drive at a safe speed and to drive through a heavily congested area. The results revealed increased alpha-band related activity (corresponding to decreased neural activity) in the right lateral PFC, including the DLPFC during the fast driving task. The participants who drove the fastest and displayed the greatest risk-taking behaviour produced the least neural activity (i.e. the highest alpha related activity). The authors attributed this to the role of the PFC in controlling risk-taking behaviour through inhibition. During the excessively fast driving task the participants were tasked with engaging in an inherently dangerous act that made controlling risk redundant. As a result they became disinhibited and the risk-controlling functions of the PFC diminished. This fits with the findings of Calhoun et al. (2000) and Peres et al. (2000) who found that neural activity in frontal areas decreased with increased speed. They also attributed this to disinhibition as higher speeds

4.2.3.2 Brain Stimulation

Beeli, Koeneke, Gasser and Jancke (2008) tested the role of the DLPFC in driving. They applied transcranial direct current stimulation (tDCS) to the left and right DLPFC of participants while they were driving in a driving simulator.

Participants drove around a 3km long simulated driving route consisting of a combination of rural and urban areas. Their results revealed that excitation of both the left and right DLFC, through the application of anodal tDCS, led to an increase in “careful” driving style without the participants noting a change in behaviour. This effect was significant for anodal tDCS but not cathodal tDCS, with anodal tDCS leading to reductions in speed and speeding violations, as well as an increase in the distance to the vehicle in front. This is consistent with the idea that increasing the neural excitability of both sides of the DLPFC resulted in safer driving, while reducing the neural excitability had no effect. The authors argued that the results might have been due to the anodal stimulation of the DLPFC reducing risk-taking behaviour. They cited the fact that the PFC has been implicated in controlling risk-taking behaviour (Bechara, Tranel and Damasio, 2000) and that tDCS to the DLPFC can alter risk-taking behaviour (Feceau et al. 2007). These findings indicate a possible role for the DLPFC in safe driving behaviour.

Sakai et al. (2014) also investigated the role of the frontal lobes in driving through the use of tDCS. They administered cathodal and anodal tDCS to the left and right DLPFC in different combinations across three separate testing sessions. A sham stimulation was used as a control condition. The participants were given a driving simulator task in which they had to maintain a constant distance to the vehicle in front that was driving along a winding road at a constant speed. The participant’s driving performance was judged in terms of the distance they kept from the vehicle in front as well as their lateral lane position (i.e. how well they kept in their lane). The results showed that driving performance was significantly better (i.e. better driving distance and lateral lane position) for right anodal/left cathodal compared to the right cathodal/left anodal and sham simulation conditions. The authors stated that the results showed that up-regulation of the left DLPFC accompanied by down-regulation of the right DLPFC resulted in improvements in fundamental vehicle control. They argued that the results indicated the dominance of the right DLFPC in driving that might have been related to anodal tDCS to the right DLPFC improving sustained attention during the driving task. The right DLPFC has been shown to be involved in sustained attention (Pardo, Fox and Raichle, 1991).

The lateralisation findings of Sakai et al. (2014) are in line with the findings of some other driving studies (e.g. Uchimaya et al., 2012). However they do contradict the findings of other studies that did not find activation of the right DLPFC during simulated driving (e.g. Calhoun et al. 2002; Just et al. 2008). Sakai et al. (2014) argued that both their study and the study by Uchimaya et al (2012) required sustained attention in order to maintain a constant distance to the vehicle in front, whereas the other studies utilised normal driving and as a result may not have required the same levels of sustained attention. This is further complicated by the fact that Just et al. (2008) found activation of the left DLPFC during a lane-keeping task and Uchimaya et al. (2003) failed to find right DLPFC activity despite using a similar driving task as Uchimaya et al. (2012). It should also be pointed out Beeli et al. (2008) failed to find any evidence of lateralisation when using tDCS. All of the contradictory findings regarding the location and lateralisation of neural activity during driving make it difficult to come to any firm conclusions. Jancke et al. (2008) found that activity was reduced in left lateral PFC when driving at high speeds, which they argued was due to disinhibition when driving at very high speeds.

It is possible that the significantly different methodologies used across driving studies result in these discrepancies. Studies investigating what parts of the brain are involved in driving have employed different driving simulation environments (e.g. some have set up simulators in MRI scanners, others have used dedicated driving simulators), different driving controls (e.g. steering responses, pedal responses, lever responses), different driving tasks (e.g. following a car in front, following a set driving route) as well as different methods to infer neural activity with (fMRI, PET, fNIR, tDCS). These differences may explain the multitude of different findings regarding the neural correlates of driving. Ultimately driving is a complex and multifaceted task, which recruits multiple areas of the brain, and the areas recruited will depend on the driving task being used and what it requires the participants to do. For example, the neural systems activated when lying down in an fMRI machine while controlling a car through a game controller may well be different from those activated while driving in a dedicated driving simulator, controlling the car with a steering wheel and pedals.

4.2.3.3 Driver Training

Evidence for the frontal lobes playing an important role in driving comes from the finding that driving performance can be improved through the training of frontal lobe related abilities and skills. Isler et al. (2008) investigated if frontal lobe executive functions are associated with driving performance. They also assessed if "higher-level", executive functioning, training had an effect on subsequent driving performance. The higher level driving training consisted of driving related skills such as hazard anticipation, risk management, emotion regulation and impulse control. The training consisted of commentary driving training (in order to improve hazard perception, situation awareness and risk management), video-based HP training (in which the participants had to detect and then select appropriate responses to on-road hazards), on-road driving self-evaluations (in which the participants compared their evaluation of their driving performance with the evaluation of a driving instructor in order to give them a better understanding of their driving ability), as well as focus groups (in which a highly qualified driving trainer discussed dangerous traffic situations with the participants in order to provide the participants with a greater awareness and insight into what constitutes a danger on the road). Driving performance was assessed via an on-road driving assessment (which examined visual search, speed control and general vehicle control and safety), a video-based HP test as well as self-report questionnaires.

Isler et al. (2008) studied 36 teenage drivers (16 and 17 year-olds) as part of a two week driver "training camp". Their findings revealed that better driving performance was related to higher levels of executive function, namely, working memory and cognitive switching. Higher general ability and a greater ability to sustain attention were also associated with higher levels of driving performance. The study revealed that higher-level driving skills training improved direction control, speed choice and visual search. The number of hazards detected and the actions towards hazards were also found to have increased. Higher-level driving skills training also resulted in a significant reduction in confidence levels for driving (overconfidence in young drivers is thought to be a contributing factor to their high accident liability e.g. Matthews and Moran, 1986), along with improvement to attitudes toward speeding, overtaking and close following (which are potential sources of road accidents).

The work of Isler et al. (2008) supports the idea that there is potentially a link between driving performance and the frontal lobes, as well as demonstrating that driving skills can be improved by training cognitive functions and skills that are dependent on frontal lobe executive functions.

4.2.3.4 Elderly Drivers

A further way to investigate the role of the frontal lobes in driving comes from testing older drivers. Research has shown that frontal cognitive functions such as executive functions deteriorate with old age (Weinberger et al. 2005). Drivers over the age of 65 have also been found to have the highest crash rates besides young drivers, with an increase in fatality rates (ECMT, 2006). Research indicates that older drivers do not necessarily pose a greater crash risk in absolute terms, as they cause fewer crashes than younger age groups per number of drivers (Ryan, Legge and Rosman, 1998). Instead they pose a greater crash risk in terms of relative crash risk, the number of crashes per distance driven (Langford et al., 2008). Testing older drivers therefore presents a way to see if these frontal cognitive functions are important to driving, by examining if the decline in these frontal functions is related to the increased crash and fatality rates of older drivers.

Richardson and Marotolli (2003) tested older drivers (aged at least 72) in order to investigate what aspects of cognition were associated with the crash risk of older drivers. In order to test cognitive abilities, the participants completed a wide range of neuropsychological tasks that measured executive functions, attention, language, verbal and visual memory, as well as visuo-motor and visuo-spatial abilities. Their driving was evaluated with an on-road test containing urban, suburban and highway driving, as well as parking manoeuvres. The results revealed that poor executive functions; visuo-motor abilities and memory were associated with poorer driving performance and increased crash risk.

The deterioration of executive functions has been proposed to be the cause of the age-related decline in many cognitive abilities, know as the frontal lobe hypothesis (Moscovitch and Wincour, 1995; West, 1996). According to the frontal lobe hypothesis, the decline in cognitive abilities and driving performance related to aging is due to a decline in executive functions. Adrian et al. (2011) explored this hypothesis by testing the executive functions and driving of drivers over the age of 60. Cognitive tests were used to test inhibition, set shifting, updating, speed of information processing as well as visuospatial abilities. Driving performance was tested on the basis of lateral position on the road, following distance, speed control, visual behaviour and communication, obeying traffic signals, mechanical operations, defensive behavior turning left, joining the traffic stream and roundabouts. The driving route contained urban and more rural areas.

The results of the study revealed that a link between executive function and driving performance, with poor driving performance associated with low scores on the cognitive tests of set shifting and updating. The authors argued that set shifting is necessary for safe driving as it allows drivers to switch between different activities to deal with the dynamic nature of traffic situations. Meanwhile updating is vital for updating working memory in order to remove unnecessary information that may occupy it. Without properly updating working memory, irrelevant information may be retained which will limit the processing of relevant information.

Daigneault, Joly and Frigon (2002) compared the executive functions and driving habits of older drivers (65 years or older) who were either accident-free or who had been involved in three or more accidents in the past five years. Executive functions were tested using Stroop, London Tower and Wisconsin Card Sorting tasks. The authors reported that the drivers with a history of traffic accidents performed more poorly on the tests of executive functions, reported safer driving behaviour on the road and had the intention of avoiding risk-taking behaviour while driving. The authors concluded that certain groups of older drivers have cognitive problems that cannot be compensated by apparently more careful driving behaviour.

Further evidence for the role of the frontal lobes in driving, and in particular the role of frontal cognitive functions such as executive function, comes from young drivers (namely those under the age of 25). Just as research has shown that frontal cognitive functions begin to decline after a certain age, research has also shown that frontal cognitive functions are not fully developed until early adulthood (Weinberger et al., 2005). Adolescence and young adulthood (when most people begin to drive) is a period of increased risk-taking and impulsiveness.

This is linked to the fact that young drivers are hugely overrepresented in road crash and fatality statistics, and display greater risk-taking behaviour when driving. It is possible that the high crash and fatality rates of young drivers are in part due to them not having fully developed frontal functions. A further parallel is that the frontal lobes themselves are not fully developed until approximately 25 years of age (Sowell, Thompson and Holmes et al., 1999). These changes in physiology may lead to the changes in frontal lobe functions, which may in turn to lead to high crash and fatality rates.

4.3 – Functional Near Infrared Spectroscopy

4.3.1 – General Introduction

In order to fully assess the role of the frontal lobes in driving, neuroimaging is needed in conjunction with realistic driving simulation. The choice of neuroimaging methodology is complicated by the fact that the most realistic simulation of a driving environment requires participants to be in a driving simulator. As a result, methods such as MEG and EEG are limited due to the electrical fields generated by the simulation equipment. These methods are also limited with regards to spatial resolution. It is possible to set up a low-fidelity driving task in an MRI scanner and as mentioned earlier, a number of researchers have carried out studies utilising this method (e.g. Graydon et al., 2004; Horikawa et al., 2005; Calhoun and Pearlson, 2012).

However there are a number of drawbacks of using fMRI in driving research. Apart from being time consuming and difficult compared to other methods, the standard of the driving simulation environment is ultimately not as realistic as those found in dedicated driving simulators. One problem is that the environment limits the use of realistic driving controls (pedals, steering wheel, gear lever etc.). For example, Horikawa et al. (2005) employed a joystick control while Graydon et al. (2004) used button controls. Even when more realistic driving controls are used other problems remain. Using fMRI requires the participants to lay down on their backs while engaging in a driving task and can also results in a restricted field of view (unless special goggles are used). fMRI is also very sensitive to movement, which places restrictions on the ability to simulate driving in as realistic manner as possible as driving does involve a fair amount of movement, including head movement. As driving involves a large amount of head movement in order to scan the environment and to use the rear and wing-mirrors, this is a significant limitation of driving studies using fMRI. As a result there is a need to investigate other methods for recording brain function during driving.

fNIR is a potentially promising neuroimaging method for driving research as it is a non-invasive, small and portable piece of equipment that can be easily utilised in a driving simulator as well as in a real car on public roads. As a result fNIR provides the ability to test frontal lobe function in realistic driving environments. Therefore fNIR may have important application for driving-related research.

4.3.2 – Fundamentals of fNIR

Functional Near Infrared Spectroscopy (fNIR) is a non-invasive neuroimaging method that used the principles of near-infrared spectroscopy (NIRs) to record functional activity in the brain. fNIR uses light in the near-infrared region (600 to 1000nm) of the electromagnetic (E.M.) spectrum to calculate levels of oxygenated ($Hb0_2$) and deoxygenated (Hb) haemoglobin (in turn providing a measure of neural activity). All biological tissue is, to varying extents, permeable to E.M. radiation of different frequencies and intensities. Different molecules scatter and absorb light of different wavelengths to differing degrees. As a result of this, different molecules produce a spectroscopic signature as a result of E.M. radiation being attenuated as it travels through biological tissue. The primary chromophores that absorb light in the near-infrared range are Hb and HbO_2. Importantly, Hb and HbO2 absorb light at slightly different regions of the E.M. spectrum, producing their own specific spectroscopic signatures. This allows for the detection of the concentrations of Hb and HbO_2 using optical spectroscopic methods.

fNIR uses light at two carefully selected wavelengths to calculate the levels of Hb and HbO_2. The near-infrared region is used for fNIR as light is predominantly absorbed by Hb and HbO_2 in the 700nm to 900nm range. The two wavelengths are selected to be above and below the isosbestic point (the point at which light is absorbed to a similar degree by Hb and HbO_2), which is 805nm (Cope and Delphy, 1988). The lower of the two wavelengths is used to record levels of Hb while the higher of the two is used to record levels of HbO_2. Depending on the methodology used, fNIR can record changes in Hb and HbO_2 or absolute concentrations. These values can then be used to calculate other measures such as blood volume, blood oxygenation and the total haemoglobin level. All of this can be done as participants take part in psychological tasks, allowing for the measurement of haemodynamic activity associated with certain psychological functions.

In order to detect haemodynamic changes, fNIR uses a modified Beer-Lambert law (Cope and Delpy, 1988; Cope, 1991). The original Beer-Lambert law relates the attenuation of light to the properties of the material that the light is travelling through, modelling ballistic photon propagation through an absorbing medium. The law describes a logarithmic ratio between the radiation falling upon a material and the radiation that is transmitted through that material. It states that for an absorbing compound dissolved in a non-absorbing medium, the attenuation (*A*) is proportional to the concentration of the compound in the solution (*c*) and the optical path-length (*d*). The law can be written as follows.

$$A = -\log(\mathrm{I}/\mathrm{I_o}) = \varepsilon\, c\, d$$

Where A is the attenuation of light (also known as the optical density or the absorbance), I_o is the light intensity incident on the medium, I is the light intensity transmitted through the medium, ε is the molar extinction coefficient of the absorbing compound (measured in micro molar per centimetre), and c is the concentration of the absorbing compound (measured in micro molar), and d is the distance between the points where the light enter and leaves the medium. Multiplying ε and c produces the absorption coefficient of the medium, μ_a.

Whereas photons travelling through non-organic matter propagate in a ballistic fashion, photons travelling through tissue are highly scattered. This process of scattering increases the path-lengths that the photons travel along through the tissue, resulting in the need to modify the law. The Beer-Lambert law only applies to non-scattering media and as a result cannot be applied to biological tissue. The modified Beer-Lambert law (Cope and Delphy, 1988; Delphy et al. 1988) takes the Beer-Lambert law and relates it to light travelling through biological tissues such as blood. This law is able to relate changes in light intensity to changes in relative concentrations of Hb and HbO_2. The modification involves introducing an additive term, G, to provide a measure of the scattering losses as well as a multiplier to account for the increased photon path lengths. The true optical distance is called the differential path-length (DP) and the scaling factor is called the differential path-length factor (DPF). The DP is calculated as follows:

$$\mathrm{DP} = DPF \times d$$

Where d is the geometrical distance between where the light enters and leaves the medium. Combining these factors creates the modified Beer-Lambert law, which can be written as follows:

$$A = -\log(\mathrm{I}/\mathrm{I_o}) = \varepsilon\, c\, d\, x\, DPF + G$$

In the modified Beer-Lambert law I_o is the light incident intensity from the light source before it passes through the tissue, I is the light intensity recorded at the detector after it has passed through the tissue, d is the distance between the light source and the detector, and ε and c are the absorptivity and concentrations of the chromophores (e.g. Hb and HbO_2) respectively. The scattering factor, G, is unknown (it is dependent upon knowing the measurement geometry and the reduced scattering coefficient, μ_s') meaning that it is not possible to derive absolute concentrations. However, as long as G stays constant throughout the measurement period, any change in the concentration of the chromophores (c_2-c_1) can be calculated from any changes in attenuation of the light (A_2–A_1). This is achieved by taking measurements at two time points. Assuming that the change in scattering is small in comparison to the change in absorption, G can be removed from the calculations determining c change over times as it can be assumed to be time-invariant:

$$(A_2 - A_1) = -\log(I_2/I_1) = \varepsilon\,(c_2 - c_1)\, d \times DPF$$

This equation allows the calculation of the differential attenuation. In order to quantify the change in the concentration and to calculate absolute concentrations is it necessary to calculate the true optical path length (by calculating *d* and *DPF*). While *d* is simple to calculate, the DPF is dependent on age, gender, the wavelengths used and also varies between subjects and tissues (Duncan et al., 1995). Average values for the DPF for an adult's head have been published (e.g. Duncan et al. 1995), however there is a significant (15%) variation between subjects. This would lead to a 15% error when determining concentration changes for individual participants. However, if the true optical path length can be determined, it is possible to calculate absolute concentrations of Hb and HbO_2, and some forms of fNIR (e.g. time-domain fNIR systems) do allow for this. Although the head is inhomogeneous, this inhomogeneity remains constant and is accounted for by G, which is cancelled out for measuring changes in the concentration of Hb and HbO_2. However, the concentration changes in Hb and HbO_2 are not homogenous as they occur in the brain but not in the other tissues that the light travels through (skin, skull and cerebrospinal fluid). As a result the Modified Beer-Lambert law underestimates changes in Hb and HbO_2 (Scholkmann et al. 2013). It is possible to correct this error by factoring in partial differential pathlengths (Hiraoka et al., 1993).

In order to determine the contribution of the different chromophores, in this case Hb and HbO_2, to the attenuation of the light, measurements must be taken at two wavelengths (one which is primarily absorbed by Hb and the other which is primarily absorbed by HbO_2):

$$\Delta A_{\lambda 1} = (\varepsilon Hb_{\lambda 1}[\Delta Hb] + \varepsilon HbO_{\lambda 1}[\Delta HbO]) \times d \times DPF$$

$$\Delta A_{\lambda 2} = (\varepsilon Hb_{\lambda 2}[\Delta Hb] + \varepsilon HbO_{\lambda 2}[\Delta HbO]) \times d \times DPF$$

With $_{\lambda 1}$ and $_{\lambda 2}$ representing the two wavelengths, ΔA representing the change in the attenuation of light at that wavelength, εHb and εHbO_2 being the extinction coefficients of Hb and HbO_2, with ΔHb and ΔHbO_2 being the changes in the concentrations of Hb and HbO_2.

fNIR exploits the mechanism of neurovascular coupling (Villringer and Dirnagl, 1995) to record localised haemodynamic activity and from that infer neural activity. As neurons do not store glucose or oxygen, neural activity is associated with an increase in cerebral blood flow (CBF) in order to provide the active neurons with the glucose and oxygen they require. There is an initial increase in the levels of Hb and a reduction in the levels of HbO_2 due to a rapid increase in the cerebral metabolic rate of oxygen ($CMRO_2$) that takes place before the increase in CBF (Malonek and Grinvald, 1996). The subsequent increase in CBF is greater than the increase in $CMRO_2$, resulting in an increase in the concentration of HbO_2 and a reduction in the concentration of Hb (Fox and Raichle, 1986).

4.3.3 Comparing fNIR to Other Neuroimaging Methods

Just as fNIR exploits neurovascular coupling to record haemodynamic activity, neurovascular coupling is what the Blood-Oxygen-Level Dependent (BOLD) contrast used in fMRI is based upon. Some studies have actually compared fNIR to fMRI in order to see if the records correlate with one another. Strangman et al. (2002) carried out a direct comparison between fNIR and fMRI during functional brain activation. Simultaneous fNIR and fMRI recordings were taken as participants engaged in a four-finger flexion/extension task. The change in the BOLD signal was compared to changes in Hb, HbO_2 and total haemoglobin concentrations. The results revealed strong correlations between the changes in the fMRI signal and changes in all of the fNIR measures. The strongest correlation was found between changes in the fMRI signal and changes in HbO_2. The authors stated that this might have partly been due to there being a superior signal-to-noise ratio for HbO_2 compared to Hb (in terms of optical measurements).

Cui et al. (2011) also carried out a comparison between fNIR and fMRI, this time using multiple cognitive tasks. fNIR and fMRI recordings were taken simultaneously during the following task: left finger tapping, go/no-go, judgment of line orientation, and an N-back working memory task that used visuospatial stimuli. fNIR recordings were taken from frontal and parietal areas. The results revealed that fNIR signals were highly correlated with the fMRI measures across all of the cognitive tasks. fNIR was found to have a significantly poorer signal-to-noise ratio. The distance between the skin and the brain, as well as the signal-to-noise ratio affected the variability of fNIR/fMRI signal correlation, with a greater distance resulting in a reduced correlation. The authors also modelled different photon path lengths and found that photon paths forming an ellipse between the emitter and the detector resulted in the largest correlations between the fNIR and fMRI signals. In line with the findings of Strangman et al. (2002), HbO_2 was better correlated with the BOLD signal than Hb.

As mentioned in the general introduction, the use of the most prominent system for measuring haemodynamic activity in the brain, fMRI, is limited in driving research due to the restrictions in environment and movement associated with the methodology. The research from Strangman et al. (2002) and Cui et al. (2011) comparing fMRI and fNIR suggests that fNIR can provide a cheaper and more flexible means with which to record haemodynamic activity during driving in both high-fidelity driving simulators and in real cars driving on the road. There is however another alternative neuroimaging method that can be used for driving research, EEG. Relatively small and mobile EEG kits are available that can be used for driving research in a number of environments.

Unlike haemodynamic measures such as fNIR and fMRI, EEG provides direct measurements of the activity of neurons in the brain. This is achieved by measuring changes in the voltage levels (electric potential) of the brain, which result from changes in the post-synaptic current flow along the dendrites of neurons as they activated or inhibited (Niedermeyer and da Silva, 2004). By averaging EEG recordings that are time-locked to specific cognitive, sensory or motor events, event-related potentials (ERPs) can be recorded which provide detailed measurements of how the brain functions in relation to these events (Luck, 2005).

Compared to fNIR, EEG does provide some advantages. A major advantage of EEG is its excellent temporal resolution, which is in the order of 1 millisecond (Gevins, Leong, Smith, Le and Du, 1995). Also, whereas fNIR is limited by the intrinsic delay in the haemodynamic response to events of interest, EEG is able to record the near instantaneous electrical activity of neurons in response to an event of interest (Niedermeyer and da Silva, 2004). This provides advantages for studies in which stimuli and responses may only occur for brief periods of time and highly accurate timings of neural responses are required. For example, Wester, Böcker, Volkerts, Verster and Kenemans (2008) used EEG as participants drove in a simulator and completed secondary visual and auditory tasks. They required accurate timings as they needed to record ERPs associated with the primary driving task and the secondary oddball tasks. However, there are areas in which fNIR can provide advantages over EEG. A drawback to EEG in relation to a method such as fNIR is its poor spatial resolution, which is in the region of several cm (Srinivasan, 1999), and the difficulty in determining the specific locations at which electric potential changes occurred. The limited spatial resolution of EEG stems from the fact that different tissues in the skull differ in terms of their conductivity (Srinivasan, 1999). As a result of this the electric potentials that pass through certain parts of the skull become distorted and tracing their origin becomes extremely difficult, usually requiring the use of highly complex and realistic head models (Srinivasan, 1999). This limited spatial resolution would limit the use of EEG in studies that need a high level of spatial resolution in order to determine which areas of the brain are being utilised during a driving task. For example, Calhoun et al. (2002) needed a high spatial resolution in order to identify the distributed networks in the brain that were associated with different aspects of driving. This would not have been possible with EEG. Accurately localising the source of neural activity is also easier with fNIR than EEG, as haemodynamic activity can be overlaid onto structural MRI scans with the help of relatively simple software (e.g. Kojima, Tsunashima and Shiozawa, 2005). One solution would be use fNIR in combination with EEG, which would allow for measurements of both the haemodynamic and direct neural activity associated with driving.

Overall, fNIR provides driving researchers with a cheaper and more flexible method or recording haemodynamic activity in the brain during driving. fNIR can in turn be combined with direct measures of neural activity such as EEG in order to provide a detailed picture of brain activity during driving. The combination of the spatial resolution and haemodynamic measures of fNIR and the high temporal resolution and electrical recordings of EEG could provide researchers with a detailed understanding of how the brain functions during driving. However, in order to achieve this, fNIR must first be further tested and validated. The number of researchers using fNIR to study the brain is growing rapidly, along with a growing amount of research literature and tools (such as software programmes for data analysis and visualisation). It is therefore important to further explore the use of fNIR for driving research in order to establish if it is a viable neuroimaging method for this area of research.

4.3.4 Types of fNIR Devices

Although there are a large number of different fNIR devices, they share some fundamental similarities in their design. The main part of an fNIR device is the sensor pad, which is placed onto the head of the participants. This part contains the light sources (e.g. LEDs) and the light detectors. Light is emitted from the light sources and is assumed to follows a 'banana' shaped path to the detectors as it is reflected back by the tissue (Okada et al., 1997). Photons emitted from a light source follow this path through tissue back to a detector located on the same approximate plane as the source. Most of the light is scattered or absorbed but a certain amount reaches the sensors. The light that reaches the sensors is encoded with the spectroscopic signatures of the molecules it has passed through. By choosing the right wavelengths of light, it is possible to capture the spectroscopic signatures of Hb and HbO_2. As with other neuroimaging methods, fNIR is susceptible to noise, particularly from the physiological functions of the body (e.g. respiration). As a result of this, fNIR devices provide automatic noise cancelling such as independent and principal component analysis, and optimal filtering (Izzetoglu et al. 2005; Ayaz et al., 2011). The sensor pad is connected to a computer, which analyses and stores the data collected by the sensor pad.

The number of light sources and detectors determines the number of channels an fNIR device has (i.e. the number of measurement locations). Each distance between a light source and a detector corresponds to a recording channel (or an optode). For example, an fNIR device with one light source and four sensors would have 4 recording channels. The area between the light sources and the detectors are the areas that haemodynamic activity is recorded from (also known as the measurement volume). The depth of this measurement volume from the surface (i.e. the skin) is a function of the distance between the light source and the detector, corresponding to roughly half the distance between the light source and the detector (Gefen, Ayaz and Onaral, 2014). As a result, the further the light source and the sensor are from one another, the deeper measurements can be taken from. The number of light sources, detectors and channels is dependent on the specific fNIR device, with some devices having 16 channels and some having as many as 64 (Ferrari and Quaresima, 2012). fNIR devices with smaller number of channels are used to record activity in single areas of the brain (e.g. the frontal lobes), while devices with a larger number of channels can record neural activity from multiple areas of the brain (e.g. the frontal, parietal lobes and occipital lobes). The sampling rate of fNIR ranges from 100Hz to less than 10Hz, while spatial resolution is in the region of 1cm (Ferrari and Quaresima, 2012).

The methods described so far in this chapter largely relate to the most commonly used form of fNIR, continuous wave (CW) fNIR. However, a number fNIR devices have been developed over the years that utilise slightly different technologies and allow for different measurements (see Elwell and Cooper (2011) for an in-depth review on this subject). These devices can be split into one of several general categories depending on the technologies they use: continuous wave, frequency-domain and time-domain. (Other near-infrared spectroscopy methods exist, such as spatially resolved and diffuse correlation spectroscopy. However, these have not yet been used extensively for functional imaging and have instead focused on tissue oximetry.)

4.3.4.1 Continuous Wave Spectroscopy

CW spectroscopy uses light sources (usually LEDs) at two wavelengths with a constant amplitude and frequency. CW devices are not capable of determining photon path lengths. As a result they are not able to measure absolute concentrations in Hb and HbO_2, only relative changes. It is sometimes possible to calculate absolute concentrations by simulating (through a Monte Carlo simulation) or assuming the photon path-lengths (Scholkmann et al., 2014). From the changes in Hb and HbO_2 it is possible to calculate other values such as the change in blood volume (BF) and the change in blood oxygenation (HbO_2 – Hb). The sampling rate and of these devices can be as high as 100Hz, while their depth penetration is limited to a few cm (Ferrari and Quaresima, 2012). Due to its relative simplicity, low cost and portability, CW is the most commonly used form of fNIR. However the cost is largely related to the number of channels, meaning that costs can increase significantly for devices with a large number of channels.

4.3.4.2 Frequency Domain fNIR Spectroscopy

Frequency domain (FD) spectroscopy also light at two wavelengths, however the light is produced by lasers and the amplitude of the light is modulated by a frequency of 100MHz or more (Torricelli et al., 2014). Changes in the amplitude and phase of the light recorded at the detectors provide direct measurements of the absorption coefficients at the two wavelengths, as well as the scattering coefficient. This removes the need to know the photon path-lengths, which allows for the calculation of the absolute values of Hb and HbO_2 concentrations. This in turn allows for the calculation of values such as the absolute tissue oxygen saturation (SO_2) and total haemoglobin concentration (HbT). FD fNIR devices also have the benefit of greater penetration depth for the light sources used (Ferrari and Quaresima, 2012). There are however a number of drawbacks to this methodology. Their sampling rate/temporal resolution is lower than that of CW devices (Ferrari and Quaresima, 2012) and they have a less favourable signal-to-noise ratio than CW devices (Elwell and Cooper 2011). The significantly more complex nature of these systems also makes them much more expensive and less portable (Ferrari and Quaresima, 2012).

4.3.4.3 Time Domain Spectroscopy

Time domain (TD) spectroscopy is a time of flight method that uses a light source (usually a laser) to provide light pulses with duration of a few picoseconds. Due to these extremely fast pulses of light, detection equipment with a temporal resolution in the sub-nanosecond scale is required (Torricelli et al., 2014). TD spectroscopy relies on the measurement of the photon distribution time-of-flight in a diffuse medium (e.g. human tissue) using these extremely fast pulses of light. By comparing the time taken for the light to travel through the tissue with a reference time from the speed of the light it is possible to calculate the absorption and scattering coefficients. This allows for the calculation of absolute values of

Hb and HbO_2 concentrations. As with FD spectroscopy, TD allows for greater depth sensitivity (Ferrari and Quaresima, 2012). However the drawbacks of a lower sampling rate/temporal resolution (Ferrari and Quaresima, 2012) and inferior signal-to-noise ratio are present (Elwell and Cooper 2011). The high-speed emitters and detectors make TD systems the most expensive and complex available (Ferrari and Quaresima, 2012).

4.3.5 fNIR Device Used in Thesis

This section will focus on the technical aspects of the continuous wave fNIR device used in this thesis, which is an fNIR Imager 100 continuous wave device (fNIR Devices, Potomac, MD, USA).

The fNIR device consists of a sensor pad fitted with 4 LED lights and 10 light sensors (Figure 2.1). The LED lights have a dual wavelength of 730nm and 850nm and the light sensors consists of silicon photodiodes with an integrated trans-impedance preamp. The arrangement of the light sources and the sensors results in data being gathered at 16 recording channels (or optodes). The sampling rate of the device is 2Hz, with a complete scan of all 16 optodes completed twice in each second (at both wavelengths). The LEDs activate sequentially (from left to right), with each of the four sensors surrounding the LED sampled in order to measure the light that has travelled from the detectors through the tissue in the head before reaching the detectors. The distance between the light source and the sensors is 2.5cm, meaning that the device is capable of recording haemodynamic activity at a depth of approximately 1.25cm, with a spatial resolution of 2-3cm^2 (Gefen et al. 2014). The sensor pad is placed onto the forehead and is secured in place with a Velcro strap. The sensor pad is connected to a control box, which contains. That is in turn connected to a computer that has a software programme on it for operating the fNIR device and storing the data collected from it. The software programme (Cobi Studio) provides an interface through which it is possible to start and stop the device from recording, take baselines, as well as placing time stamps (called markers) into the data to note a point in time in which something important took place (e.g. a stimulus appeared).

The photons produced by the LEDs follow a characteristic banana shaped path through the target tissue and then back to a detector on the same approximate plane (Figure 4.2). Most of the light transmitted from the light sources is absorbed or scattered but some makes its way to the detector sites. The attenuation of the photon density at the two wavelengths (as recorded at the detectors) provides the information (the spectroscopic signature) required to calculate the relative concentrations of Hb and HbO_2.

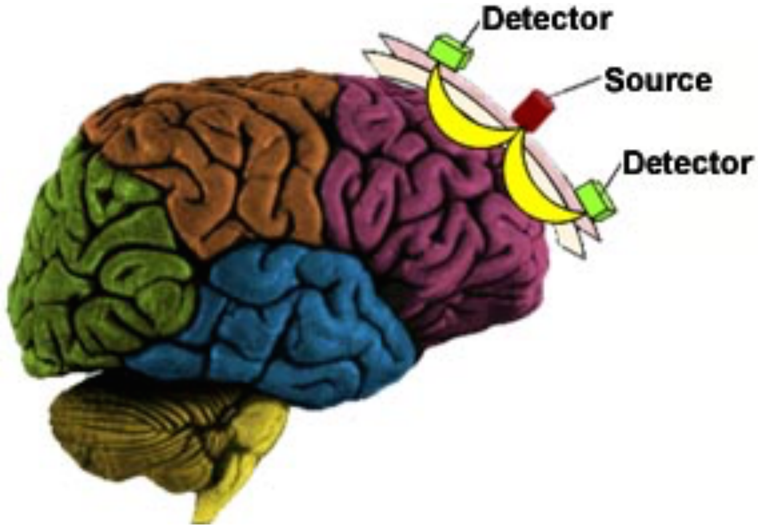


Figure 4.1. Typical pathlengths taken by photons as they travel from an LED light source, through brain tissue and on to a light detector.

For this process to work, the sensor pad on this system needs to be in contact with the skin. As a result this fNIR device is primarily used to monitor hemodynamic activity in the frontal lobes, namely the DLPFC as placing the sensor on a participant's forehead accesses these areas. The DLPFC is the most accessible area of the frontal lobes when using fNIR as it is not obstructed by the eyebrows and the eyes. As mentioned above, this fNIR device is capable of accessing regions 1-2cm deep, with a spatial resolution of 2-3cm^2. As a result, this fNIR device is especially suitable for recording activity in the frontal lobes as they are a relatively short depth below the skin on the forehead (due to the shape of the brain) and the skull is relatively thin in this region. This allows the light from the LED emitters to better reach the cortex.

Before undergoing any analysis, the data are passed through a low-pass filter to remove physiological noise in the signal, including the signals from respiration and the cardiac cycle (Izzetoglu et al. 2005). As detailed earlier, to calculate the changes in Hb and HbO_2, fNIR uses a modified Beer-Lambert Law. The combined changes in the intensity of the light received at the sensors (at 730nm and 850nm) are used to calculate the changes in Hb and HbO_2 over time (absolute concentrations cannot be calculated with this device). This process first involves creating baseline readings for the relative concentrations of Hb and HbO_2. The baseline also serves to factor in individual differences in the properties of participant's skin, the shape of their heads and the shape of their brains. Further readings are then taken at different time points to calculate the change in Hb and HbO_2 over time (e.g. during psychological tasks). Calculating absolute concentrations of Hb and HbO_2 is not possible with this device. These changes in Hb and HbO_2 can then be used to provide other measures of haemodynamic activity such as the change in blood volume and the change in blood oxygenation:

$$\Delta \text{ Blood Volume} = \Delta HbO_2 + \Delta Hb$$

$$\Delta \text{ Blood Oxygenation} = \Delta HbO_2 - \Delta Hb$$

The procedure for using this fNIR device first involves properly placing the sensor pad on the participant's forehead and securing it in place with the Velcro straps. A test must then be run to establish if the device is recording a clean signal. If there is hair in the way of the light sources or detectors, or if the sensor pad is loose in certain places and is not in contact with the skin, the recorded signal will not be an accurate measure of haemodynamic activity. The signal can be checked on the computer that runs the custom software for the device. Once any adjustments have been made and the signal is at an acceptable level a baseline reading must be taken. This involves taking 10 seconds worth of haemodynamic recordings while the participant is in a state of rest. The purpose of the baseline is to establish relative concentrations of Hb and $Hb0_2$ at rest, which can be used to measure functional changes in Hb and HbO_2 associated with psychological tasks.

Once the baseline is taken the psychological task can begin. As soon as the task begins the fNIR device is set to record any changes in haemodynamic activity relative to the baseline. The device is left recording until the end of the task when it is stopped from recording. The device is left to record any changes throughout the duration of the task. As soon as the recording is stopped, several files containing data are created. One file contains raw data, another file contains data that have already had the changes in Hb and HbO_2 over the tasks duration calculated, and another file contains information on any markers that have been used during the recording process. These values are provided for each of the 16 recording channels.

4.3.6 History of fNIR

The use of fNIR as a neuroimaging method can be traced back to the work of Glenn Millikan in the 1940's (for a review of this work see Chance, 1991), who developed the muscle oximeter. Frans Jöbsis, who was the first person to use in vivo NIRS, took the next steps in the development of optical methods for measuring changes in haemoglobin in the brain. He discovered that the high level of brain tissue transparency in the near-infrared range allows for non-invasive and real-time detection of the oxygenation of haemoglobin (Jöbsis, 1977). After carrying out experiments using NIRs on laboratory animals, NIRs was used record cerebral oxygenation in newborn babies with health problems (Brazy et al., 1985). Important work was also undertaken by Marco Ferrari who used prototype NIRs devices to measure brain oxygenation changes in adults (Ferrari et al., 1985). David Delpy, working at University College London, developed a number of prototype NIRs devices. This led to the first quantitative measurements of a number of oxygenation and hemodynamic parameters in sick newborns. This included changes in HbO_2 and Hb, total hemoglobin concentrations (tHb; $tHb=HbO_2 + Hb$), cerebral blood volume, and cerebral blood flow (Wyatt et al., 1986). This work led to the development of a 4-wavelength system (Cope and Delpy, 1988) that was the basis of the single-channel continuous wave NIRO-1000, which was the first commercial NIRs system built in 1989.

The first studies using NIRs as a functional measure of changes in neural activity came in the early 1990's just after the discovery the discovery of the BOLD signal and its potential use for functional studies of the brain (Ogawa et al. 1990). One of the first functional NIRs studies carried out (although it was delayed in being published) involved using the NIRO-1000 system to record neural activity in the occipital cortex during photic stimulation (Kato et al., 1993).

The results revealed a significant increase in HbO_2 during photic stimulation. These results were repeated by other research groups using the NIRO-500 system (e.g. Meek et al., 1995). Hoshi and Tamura (1993) used the OM-100A system to record increased HbO_2 and decreased Hb levels in the PFC during tasks that exerted mental workload. The NIRO-500 system was also used to record an increase in HbO_2 and tHB in the PFC during a mental calculation task. Colier et al. (1997) was the first to use single-channel CW fNIR to record activity in the motor cortex, showing evidence of the heart beat in the HbO_2 signal. This prototype device had a high temporal resolution (10Hz) and was able to detect oxygenation changes in the motor cortex in response hand and foot movements. In 1997 the fNIR was first used to monitor sequential activation in the brain. Hoshi and Tamura (1997) reported on sequential activation in the PFC during cognitive tasks.

All of these studies employed single-channel CW fNIR devices. During this period development also focused on the creation of multi-channel CW fNIR devices that were able to record activity over larger areas of the cortex. Major advances in multi-channel fNIR came from the Japanese companies Hitachi and Shimadzu. In 1994 Hitachi introduced a 10-channel CW device. This introduction of this device allowed for cortical mapping using fNIR. Cortical mapping involves acquiring multiple reflectance measurements at small source detector separations (3cm) over a large area of the head at the same time. Maki et al. (1995) investigated the ability of this device to record neural activity by looking at the hemodynamic responses elicited in the motor cortex during a finger-tapping task. Static topograms of the changes in in HbO_2, Hb and tHb were compared with anatomical MRI scans to reveal that the haemodynamic changes were taking place in an area of the motor cortex along the central sulcus. Thus study demonstrated the ability of fNIR to accurately map brain activity in humans. In 2000 Hitachi released a 24-channel device (the ETG-100). Hitachi then released 68-channel (ETG-700), 52-channel (ETG400) and 72-channel systems (ETG-7100) in 2002, 2003 and 2005 respectively.

In 2001, Shimadzu released a 42-channel fNIR device (OMM-2001). Miyai et al. (2001) used this system to visualize the cortical activity associated with the human gait. The results revealed that walking activities were associated with bilateral increases in levels of HbO_2 in the medial primary sensorimotor cortices and the SMA. Alternating foot movements activated similar but less broad regions, with gait imagery increased activities caudally located in the SMA. Shimadzu released a 52-channel (OMM-300) and a 64-channel (NIRStation) system in 2004.

Multi-channel fNIR devices have also been developed in Europe and the United States. TechEn in the United States has developed a number of systems, and their most recent device (CW6) is able to monitor neural signals from any combination of 32 lasers and 32 detectors. A number of systems (with different number of sources and detectors) have also been developed by NIRx Medical Technologies. Koch et al. (2010) used one of these systems to create functional mapping of the somatosensory cortex. The first multi-channel CW fNIR devices in Europe were built at the Radbound University of Nijmegen in the Netherlands, and Humboldt University in Berlin. The 12-channel device built at Nijmegen was used to study haemodynamic activity during a study on language switching and translation (Quaresima et al. 2002). The results of the study revealed changes in oxygenation in the lateral frontal cortex during language switching and translation. Broca's area was found to be involved in translation irrespective of the direction of translation. A spinoff company, Artinis Medical System, was created from the Radbound University of Nijmegen. It developed a multi-channel CW device with up to 96 channels (Oxymon MkIII). Humboldt University developed built a 22-channel CW system (Obrig and Villinger 2003).

During this same period, FD and TD fNIR systems were also under development. In vivo TD NIRs was first used in 1988 (Delpy et al. 1988) in order to directly record the average pathlengths of photons travelling through the head of a rat. The data from this study has in turn been used to calculate the DPF for other studies. TD fNIR was then used to calculate kinetics of Hb deoxygenation in cat brain during mild hypoxia (Chance et al., 1988). DPF values were then measured in newborn infants and adults (van der Zee et al., 1993). FD NIRs was first used in 1990 to measure the time-dependent migration of photons through tissue (Chance, 1990). Portable FD NIRs devices were soon developed and used to measure the optical pathlengths in the head's of newborns and adults (Duncan et al., 1995), as well as measuring HbO_2 saturation in the adult PFC (De Blasi et al., 1995). From 1995 onwards, FD NIRs has been used to detect the neuronal signal related to optical changes that is directly associated with neuronal activity (e.g. Gratton et al., 1995). This study was able to create 2D functional images of the occipital cortex during visual stimulation. Studies have also used FD NIRs to create real-time optical topographic images of the adult brain during activation of the motor cortex (Franceschini et al., 2000).

4.3.7 Research Using fNIR

The role of the frontal lobes in driving has also been investigated through the use of fNIR. Tsunashima and Yanagisawa (2009) used fNIR to examine frontal lobe activity during simulated driving. They recorded frontal lobe activity when participants drove in the simulator with and without Adaptive Cruise Control (ACC) in order to assess whether ACC reduced workload while driving. Their task was to follow a vehicle in front of them with or without ACC. During the driving tasks, the vehicle in front would stop and then starting moving again. In the condition without ACC, the participants would have to brake and accelerate as appropriate. In the condition with ACC the participants only had to steer the vehicle. They found that driving without ACC resulted in increases in HbO_2, whereas driving with ACC did not result in increases in HbO_2. They found that during the no ACC condition, the outer portions of the DLPFC (on both the left and right sides) showed significant levels of activity. Interestingly, HbO_2 values were generally higher towards the end of the task. Hb levels decreased when HbO_2 levels increased and increased when HbO_2 levels decreased.

Tsunashima and Yanagisawa (2009) also carried out a workload task (mental calculation task) in order to assess the ability of fNIR to record differences in workload. They also used a NASA-TLX (Task Load Index) workload scale (Hart and Staveland, 1988) to record subjective levels of workload. They found that HbO_2 increased with increasing task workload. Subjective workload scores also increased with increasing task workload. The fNIR data and subjective workload scores displayed a strong correlation with each other.

Kojima, Tsunashima and Shiozawa (2005) used fNIR to monitor the frontal lobe activity of people while they drove in a train simulator. The participant's were either trained or untrained in using the train simulator. The participant's task was to drive along a 2.3 km stretch of track that had three train stations located along it. The participants had to stop at each station, wait, and then continue driving. When the participants braked in order to stop at one of the stations, HbO_2 levels increased and Hb levels decreased. HbO_2 levels were at their highest when the train was arriving at a station, while Hb levels were lowest at this point. This activation was recorded on both sides of the frontal lobes. This effect was greater in the untrained driver and the greatest levels of frontal lobe activation took place when the untrained driver failed to stop at one of the train stations correctly. The authors attributed these differences to the increased task difficulty for the untrained driver and with that, the increased task workload associated with this and the braking events as a whole. Structural MRI scans were then used to map the fNIR data from the braking phases onto the frontal lobes. They showed that the DLPFC was the particular area of the frontal lobes that was being activated.

fNIR has also been employed to monitor neural activity during driving on public roads. Yoshino, Oka, Yamamoto et al. (2013a) monitored brain activity with fNIR during acceleration and deceleration tasks on a public road. The participants were required to accelerate from 50 km/h to 100 km/h in the acceleration task and decelerate from 100km/h to 50 km/h in the deceleration task. The tasks were completed in the day and at night. They calculated the correlations between measures of cerebral oxygen metabolism (Hb, cerebral oxygen exchange and the phase angle k (which is the ratio of Hb and HbO_2)) and cerebral blood flow (HbO_2 and cerebral blood volume) at specific brain sites (that had been mapped using structural MRI) with the average vehicle acceleration/deceleration in the tasks.

The majority of the correlations for the haemodynamic measures at the specific brain sites occurred during deceleration rather than acceleration. Measures of increased cerebral oxygen metabolism, while measures of cerebral blood flow decreased. This demonstrated that deceleration led to greater haemodynamic activity in the brain. The faster the deceleration the greater the increases in the measures of cerebral oxygen metabolism were. This effect was particularly apparent in the frontal eye field. The authors attributed this effect to the rapid visual processing of the environment that occurs when driving. At night, faster deceleration also led to greater decreases in the measures of cerebral blood flow across the prefrontal cortex and parietal lobes. The authors state that this was due to oxygen metabolism associated with transient ischemic changes.

These findings are in line with those of Kojima et al. (2006) who found that the greatest levels of prefrontal cortex activity occurred when the participants were braking in order to stop at a train station. Both studies show that braking (deceleration) produces the greatest levels of frontal lobe activity. It should be noted that while HbO_2 levels increased in the Kojima et al. (2006) study and decreased in the Yoshino et al. (2013a), these results are not necessarily contradictory. Research has shown that upon the presentation of a stimulus, there is an initial dip in the Blood Oxygenation Level Dependent (BOLD) signal that corresponds to transient increases in the level of Hb (Malonek and Grinvald, 1996). This is thought to reflect a rapid increase in the cerebral oxygen metabolism before the corresponding increase in cerebral blood flow. The subsequent increase in cerebral blood flow leads to an increase in the levels of HbO_2 and a reduction in the concentration of Hb (Fox and Raichle, 1986). Given the shorter braking durations in the Yoshino et al. (2013a) study, their results may largely be based around the early increases in Hb.

In a further study, Yoshino, Oka, Yamamoto et al. (2013b) used fNIR on public roads but this time they utilised a wider range of driving tasks. In this study, neural activity was monitored while the participants were: accelerating, decelerating, driving at a constant velocity and making a U-turn. Changes in cerebral oxygen exchange and cerebral blood flow were calculated. The experiment was completed during the day and again at night. The FEF (BA8) was the area most activated across the different driving tasks. The authors stated that this was due to increased voluntary eye movements (which are controlled by the FEF) in 3-dimensional space. The fact that that frontal eye field activity was lower in the constant velocity-driving task indicated that FEF function is greater during changes in speed. Acceleration led to significant increases in the change of cerebral oxygen exchange in the prefrontal cortex, namely right BA8 and left BA46 (which is a part of the DLPFC). This was attributed to the DLPFC being involved in internal monitoring in order to reach the requested speed within the allowed distance. Driving at a constant velocity only led to significant activity in the FEF (although at a lower level than in the other tasks as previously explained). The authors attributed this to the lower levels of workload required in this task as no significant handling inputs were required in this task. This is in line with Kojima et al. (2006) who reported that driving at a constant velocity did not result in significant increases in frontal lobe activity compared to braking in order to stop at a train station.

Beyond driving research, fNIR has also been utilised to study a wide range of brain functions. Tables 4.1 through 4.5 contain a list of some of the psychological research that has been undertaken using fNIR. This includes research on driving, executive functions, motor skills, memory and attention, and language.

Table 4.1. List of some studies that have used fNIR to monitor DLPFC activity during driving.

Researchers	Task	Findings
Tsunashima and Yanagisawa (2009)	Simulated Driving	Increased HbO_2 when driving without cruise control in DLPFC
Kojima et al. (2006)	Simulated Driving	Increased HbO_2 and decreased Hb during braking in DLPFC
Yoshino et al. (2013a)	Driving	Greater haemodynamic activity for deceleration compared to acceleration across PFC
Yoshino et al. (2013b)	Driving	Greater haemodynamic activity for deceleration compared to constant velocity across PFC

Table 4.2. List of some studies that have used fNIR to record DLPFC activity during executive function tasks.

Researchers	Task	Findings
Herrmann et al. (2005)	Go/No-Go Task	HbO_2 increase in inferior PFC during No Go
Boecker et al. (2007)	Stop Signal Task	Right PFC activation during inhibition
Schroeter et al. (2002)	Stroop Task	Increase in HbO_2 and reduction in Hb in superior lateral PFC during incongruent task
Kaneko et al. (2011)	Digit Span Task	Increase in HbO_2 during digit scan backwards trials
Cutini et al. (2008)	Task-Switching	Activation of DLPFC and superior frontal gyrus during switching trials

Table 4.3. List of some studies that have used fNIR to record DLPFC activity during motor activity tasks.

Researchers	Task	Findings
Miyai et al. (2001)	Walking	Increased HbO_2 in SMA and primary sensorimotor cortex bilaterally
Sato et al. (2007)	Finger tapping task	HbO_2 increase and Hb decrease in motor cortex (largest changes in contralateral hemisphere)
Holper et al. (2009)	Finger tapping task	Complexity of finger tapping affects HbO_2 level of change in primary motor cortex
Suzuki et al. (2004)	Walking	3 to 5km/h walking activated bilateral medial SMC, 9km/h activated PFC and PMC
Hatakenaka et al. (2007)	Pursuit Rotor Task	Reductions in changes in cortical oxygenation associated with learning of motor skills

Table 4.4. List of some studies that have used fNIR to record DLPFC activity during memory tasks.

Researchers	Task	Findings
Schreppel et al. (2008)	n-back Task	Activation differences in PFC depending on relevance of stimuli
Toichi et al. (2004)	Continuous Performance	Increase in HbO_2 and tHb in DLPFC
Cutini et al. (2011)	Short-Term Memory	Increase in HbO_2 in posterior parietal cortex bilaterally
Nakahachi et al. (2010)	Trail Making Test	Activation of DLPFC and VLPFC bilaterally during task F and left VLPFC for task R
Basso Moro et al. (2013)	Logical Memory	Increased oxygenation in VLPFC during encoding, with increased oxygenation in VLPFC, DLPFC during retrieval

Table 4.5. List of some studies that have used fNIR to record DLPFC activity during language tasks.

Researchers	Task	Findings
Schecklmann et al. (2008)	Verbal Fluency	Increased HbO_2 and decreased Hb in inferior DLPFC and temporal cortex
Noguchi et al. (2002)	Syntactic and Semantic Decision	Left inferior gyrus more active during Syntactic
Herrmann et al. (2006)	Verbal Fluency Task	Less activity in left DLPFC of older participants
Rossi et al. (2011)	Phonological Cues	Hb decrease in left fronto-temporal areas for phonotactically legal pseudo- words compared to illegal pseudo-words
Kovelman et al. (2008)	Picture Naming in Bilinguals	Increased activation in left posterior temporal regions when task executed bilingually
Quaresima et al. (2002)	Language Switching and Translation	Increased activity in Broca's area during translation, regardless of direction

4.5 Conclusions and Potential for Driving Research

The wide range of research undertaken using fNIR demonstrates the ability of this technique to be used in psychological research. The research listed shows that fNIR has been successfully used to test functions associated with the frontal lobes, including higher cognitive functions such as executive functions. fNIR is also increasingly being used in driving research, to test what role the frontal lobes (and the brain as a whole) play in driving. This indicates that fNIR is a suitable means with which to explore the role that the frontal lobes play in driving.

Chapter 5 - Validating fNIR

5.1 Abstract

Before using fNIR to record DLPFC activity during driving it was necessary to first validate the fNIR device in order to establish what cognitive processes the device is able to record haemodynamic changes from, as well as other factors such as the sensitivity of the device to haemodynamic changes. These results could then be used to help design appropriate driving tasks and designs to use with the device. In order to achieve this, participants completed three classic frontal tasks (the WCST, the CBT and the COWAT), during which changes in DLFPC blood oxygenation were recorded using fNIR. The results revealed significant increases in DLPFC blood oxygenation during the WCST and COWAT compared to baseline conditions. However, no significant differences were found for the CBT. The fNIR results were in line with the behavioural results for the WCST and the COWAT but not the CBT. Overall the results indicated that the fNIR device was primarily recording changes in blood oxygenation related to increases in task workload, with the possibility that levels of task inhibition were also contributing to changes in blood oxygenation.

5.2 Introduction

Before using fNIR to explore what role the frontal lobes play in driving, it is first vital to validate the fNIR device in order to establish whether or not the device is reliably recording haemodynamic changes and what psychological stimulus is leading to any changes. In order to do this, a selection of classic frontal tasks that have been shown to elicit neural activity in the DLPFC will be tested on participants while haemodynamic activity in their DLPFC is recorded using the fNIR device. As the fNIR device used in this thesis is best suited to recording activity in the DLPFC, tasks that have been found to involve the DLPFC will be selected. The tasks will be the Wisconsin Card Sorting Task (WCST), the Corsi Block Tapping task (CBT) and the Controlled Oral Word Association Test (COWAT).

The process of selecting the tasks involved an extensive literature review, focusing on well-established cognitive tasks that have been found to be dependent on areas of the PFC. The WCST, CBT and COWAT were chosen as previous literature (detailed in the introduction sections below) has shown them to reliably recruit areas within the PFC, including the DLPFC, which is the area most accessible by the fNIR device that will be employed in this study. These three tasks were also selected, as they each require a range of different frontal cognitive functions to be completed successfully. The WCST primarily tests set-shifting, whereas the CBT tests visuo-spatial working memory and the COWAT tests verbal fluency. Given that the aim of the study is to explore what the fNIR device is able to record, the decision was taken to employ tasks that will require a range of cognitive abilities. A further reason for selecting these three tasks was practical one. These three tasks are relatively easy to adapt into computerised versions that will allow for the entire testing and data collection process to be automated and completed with little interference from the experimenter. This allows for each task to last the same amount of time for each participant, which in turn aids with the fNIR analysis (as task events could more easily be time-locked to the fNIR recordings). In the case of the WCST and the CBT, fully computerised versions already exist and have been validated (Lie et al., 2006; Toepper et al., 2010). This will allow for the results of this study to be compared to those previous studies in order to help verify and interpret the results.

5.2.1 Wisconsin Card Sorting Task

The WCST (Grant and Berg, 1948) is a test of set-shifting (cognitive flexibility) and is one of most widely used neuropsychological tests in the history of clinical psychology, used to diagnose dysfunction of the frontal lobes (Retzlaff, Butler and Vanderploeg, 1992). The test is a card-matching task in which a person must match a presented stimulus card to one of four reference cards on the basis of an unknown rule (criterion). For example, the cards may be matched on the basis of the shape, colour or number of the objects on the card. The job of the participant is to use a method of trial and error to work out what the correct criterion is (shape, colour or number). The participant choses a reference card to match the stimulus card with and an experimenter says whether or not the participant is correct. The matching criterion will stay the same for a certain period of selections, after which the experimenter will change the rule. The participant then has to learn the new rule through trial and error. As a result the fundamental process required in the WCST is rule learning. Performance on the WCST is measured in a number of ways, including: preservative errors (responding on the basis of a previously ‘correct’ criteria which no longer applied despite having been informed that the selection was incorrect), set loss errors (errors occurring after three consecutively correct responses), any other errors made, as well as the number of dimensions found, the number of dimensional changes achieved (Heaton et al., 1993).

The WCST was originally developed and utilised as a measure of abstract reasoning ability and was also found to be a measure of executive functions (Luria, 1973). The WCST is thought to measure executive functions such as: abstract reasoning, problem solving, planning, organised search, response maintenance, using feedback to shift cognitive sets and change behavioural responses, directing behaviour towards achieving a goal and controlling impulsive responding (Chelune and Baer, 1986; Lezak, 2004). It was not until later that the WCST began being used as a neuropsychological test to diagnose patients with frontal lobe dysfunction (Milner, 1963), but is now considered to be a key measure in the diagnosis of frontal lobe dysfunction (Braff et al., 1991).

Patients with diverse damage to the frontal lobes have been found to perform poorly on the WCST in comparison to healthy controls (for example Axelrod et al., 1996; Janowsky et al., 1989). Barcelo and Knight (2002) used the WCST on stroke sufferers who had damage to the frontal lobes and normal controls. The results revealed that those with damage to their frontal lobes performed significantly worse than the normal group. These findings are supported by research that tested individuals with frontal lobe lesions caused by penetrative brain injuries (Grafman et al., 1990). Lesion studies have also compared WCST performance between people with frontal lesions and non-frontal lesions, with frontal lesions found to result in significantly poorer performance than non-frontal lesions (Bornstein, 1986; Eslinger and Grattan, 1993; Heaton et al., 1993). However a number of studies failed to find significant differences between those with frontal and non-frontal lesions (Anderson et al., 1991; Axelrod et al. 1996; Grafman et al., 1990), indicating that impaired WCST performance may be sensitive to but not specific to frontal lobe damage.

Neuroimaging research has also provided evidence that the frontal lobes are activated during the WCST, with activity in the DLPFC and ventromedial prefrontal cortex (VMPFC) found in a number of studies. For example, Haines et al. (1994), Kawasaki et al. (1993) and Weinberger (1986) reported activation of the DLPFC. Meanwhile Mentzel et al. (1998), Tien et al. (1998) and Volz et al. (1997) reported activation of the DLPFC and VMPFC. There is no clear evidence of laterality from neuroimaging research, with some studies finding greater activity in the left hemisphere (e.g. Haines, 1994) whereas other studies have reported greater activity in the right hemisphere (e.g. Volz, 1997). As with the finding from lesion studies, neuroimaging data indicates sensitivity for areas of the frontal lobes to the WCST but not specificity. For example, Mentzel et al. (1998) also found activity in the thalamus and basal ganglia; Nagahama et al. (1996) found activity in the striate cortex and inferior parietal cortex, while Tien et al. (1998) also reported increased activity in the inferior parietal cortex.

These neuroimaging studies suggest that the DLPFC and VMPFC are recruited during the WCST. This is supported by lesion studies that have investigated the effects of localised DLPFC lesions. As mentioned in Chapter 4, the executive functions involved in the WCST, such as task switching, are associated with the DLPFC and there is evidence that the DLPFC is a prefrontal area heavily involved in the WCST. Milner (1963) proposed that the DLPFC is important for performance on the WCST. She had patients who had undergone cortical excisions complete the WCST. The study revealed that frontal excisions had a much greater affect on task performance than posterior excisions. The results of the study also revealed that removal of the inferior and orbital sections of the frontal lobe had no effect on WCST performance, whereas all removals that encroached upon Brodmann Area 9, which includes a part of the DLPFC, resulted in poor performance on the WCST. Other neuropsychological studies on patients with DLPFC excisions provide further support to this theory (Rezai et al., 1993; Weinberger, Berman and Zec, 1986). This position is further supported by lesion studies that have reported that patients with specific damage to the DLPFC have inferior performance on the WCST in comparison to patients with lesions in other prefrontal areas such as the orbito-frontal cortex (Ettlinger, Teuber and Milner 1975; Grafman et al., 1994).

Lie et al. (2006) carried out an fMRI study on the WCST in order to understand the neural substrates utilised during the task. They used a modified design in which participants completed three WCST tasks: a standard WCST, a WCST in which the participants were informed of the sorting criteria every four trials and a baseline WCST in which the stimulus card was an exact match for one of the reference cards. The ultimate difference in these tasks regards cognitive switching, with uninstructed, instructed and no cognitive switching required for the different tasks. The behavioural results revealed that performance was best on the baseline task, followed by the informed WCST and then the uniformed WCST. Performance was measured in terms of cognitive costs (a comparative measure of reaction times for the different conditions), errors (preservative, set-loss, other errors) as well as the number of dimensions and dimensional changes achieved.

The fMRI results revealed a complex neural network underlying performance on the WCST. This network included frontoparietal regions and the striatum. The right VLPFC was found to be related to simple WM operations, while right DLPFC was related to more complex WM operations. The rostral anterior cingulate cortex and temporoparietal areas were responsible for attention and error detection. The caudal anterior cingulate cortex and the right DLPFC were associated with increased attentional control with regards to the increase of demands of working memory and cognitive control. Non-frontal neural activity was related to set-shifting (cerebellum) and working memory representations (superior parietal cortex, retrosplenium).

One important takeaway from this study is that while the WCST has been found to be sensitive to frontal lobe damage and elicits activity in frontal areas, it is not specific to frontal areas. However there is extensive evidence that performance in the WCST is reliant upon the frontal lobes, in particular the DLPFC.

5.2.2 Corsi Block Tapping Task

The CBT task (Corsi, 1972) is a neuropsychological test of visuo-spatial working memory (VSWM). The test involves copying an experimenter after they tap a series of blocks in a specific order. There are 9 blocks that are identical in appearance. In its standard form the experimenter usually begins by touching 2 blocks (encoding phase), with the participants having to reproduce the series of block taps (response phase) in the same sequence (‘forward’ task) or in reverse order (‘backward’ task). If a certain proportion of the blocks are tapped in the correct order (usually 1/2, 2/3, or 3/5 of the trials per sequence length) then the experimenter will increase the sequence length by 1 block at a time. Performance on the CBT task reduces as participants are required to remember more block locations in a sequence that is not perceived to be meaningful (Toepper et al., 2010). When the number of incorrect reproductions exceeds a certain proportion the test stops. The Corsi span averages 5 in normal adult subjects (Kessel et al., 2000). Performance on the CBT task can be measured in terms of the proportion of correct responses, as well as the reaction time of correct responses (Toepper et al., 2010).

Berch et al. (1998) described the CBT task as the most important non-verbal task in neuropsychological research. As well as being widely used in diagnostics for assessing VSWM, the CBT task also measures spatial attention (Smyth and Scholey, 1994). The CBT task is often used as part of test batteries to diagnose diseases such as Alzheimer's disease (Carlesimo et al., 1996), Korsakoff's syndrome (Haxby et al., 1983), schizophrenia (Chey et al., 2002), as well as helping to localise focal brain lesions (Milner, 1971). The CBT is also widely used in research as a measure of VSWM (Fischer, 2001; Bo et al., 2011).

VSWM is a neuropsychological concept, which underlies the ability to store, maintain and manipulate spatial information (Toepper et al., 2010). VSWM has mainly been viewed in regards to the working memory model of Baddeley and Hitch (1974,1986). This model consists of two components: the visuospatial sketchpad and the central executive. The visuospatial sketchpad is responsible for the storage of object-related and spatial information and associated with right-sided or bilateral parietal brain regions (Kessels et al., 2000). The central executive is responsible higher-level cognitive operations. These can be divided into different sub-functions (e.g. manipulation and updating information, dual task coordination, and inhibition (Collette and Van der Linden, 2002)).

Lesion studies have revealed that damage to the frontal lobes results in relatively poor performance on the CBT task. Damage to the PFC, in particular the right DLPFC, is associated with deficits in spatial span (van Asselen et al., 2006; Bor et al., 2006; Teixeira Ferreira et al., 1998). Bor et al. (2006) analysed performance on spatial span tasks (similar to the CBT) in individuals with frontal lesions. The results revealed that the individuals with frontal lobe lesions were significantly impaired on the CBT. Those with lesions to the right DLPFC were particularly impaired. van Asselen et al. (2006) had individuals with brain lesions complete the CBT. Damage to the right DLPFC and the right posterior parietal cortex was found to impair performance on the CBT. These individuals struggled to maintain spatial information over extended periods of time, indicating the importance of these areas for spatial working memory. Kessels et al. (2000) carried out a meta-analysis on the effects that lesions had on performance on spatial span tasks (similar to the CBT). They found that individuals with frontal lobe lesions performed worst on the task through having the lowest Corsi spans. Their performance was worse than that of individuals who had undergone temporal lobectomies.

Bor et al. (2001) used PET to analyse neural activity during spatial span tasks (similar to the CBT). The results revealed increased activity in the DLPFC as well as the VLPFC during the task. The authors concluded that these areas are involved spatial working memory. Owen et al. (1996) had participants complete the CBT while they underwent PET imaging. The results showed that the CBT elicited activity in the PFC, primarily in the VLPFC.

An adaptation of the CBT task is the Block Suppression Task (BST) that is thought to require spatial inhibition as a key component to the task (Beblo et al., 2004). As mentioned in Chapter 4, inhibition is an executive function that is heavily associated with areas such as the DLPFC. The BST test an individual's ability to ignore irrelevant spatial information during the CBT and only attend and respond to the target blocks in a sequence. The BST has been tested on individuals Alzheimer's disease. Toepper et al. (2008) reported that the individuals with Alzheimer's were specifically impaired on the BST compared to healthy controls. This indicated that the BST was an appropriate task for differentiating between those with Alzheimer's and those who do not have it, meaning that the BST may be a useful tool for diagnosing those with neurological disorders that affect the inhibitory functions of the PFC.

Toepper et al. (2010) carried out an fMRI study using both the CBT task and the BST to investigate the neural substrates responsible for the maintenance of short-term spatial information (for the CBT task) as well as the inhibition of spatial information (for the BST). Behavioural results revealed that correct responses were fewer and reaction times were longer for the CBT in comparison to the BST. fMRI results revealed that during the encoding phase of the CBT task there was increased activity in the right DLPFC and VLPFC. During the encoding phase of the BST there was increased activity in the left DLPFC. These findings demonstrate that encoding of spatial information and the inhibition of spatial information recruit different areas of the PFC.

The authors argued that these results supported a process-specific fractionated model of spatial working memory in the PFC. They proposed that the encoding of visuo-spatial information recruits the VLPFC and the DLPFC (as used in the CBT), whereas higher executive functions are associated with the DLPFC. The authors stated that the results suggest that the left DLPFC is involved in inhibiting spatial distraction during working memory processes. This is supported by research that has indicated that frontal activity reflects inhibitory functions that prevent distractors from interfering with working memory (Vogel and Machizawa, 2004). The fMRI results also indicated that parietal activity was correlated with working memory load. The BST therefore provides a further means of investigating PFC haemodynamic activity using fNIR.

5.2.3 Controlled Oral Word Association Task

The COWAT (Benton and Hamsher, 1976; Benton, Hamsher and Sivan, 1994) is a test of phonemic verbal fluency that measures the spontaneous production of words. The task requires participants to produce words beginning with a specific letter within a set amount of time (usually 1 minute), while avoiding any repetition. So for example, a participant may be set the task of listing as many words beginning with the letter R within the space of 1 minute. Performance on the COWAT is measured in terms of the number of correct words generated (i.e. excluding non-words and repeated words). As people have been found to produce words in clusters with pauses in between (Bousfield and Sedgewick, 1944), there is no calculation for the rate at which words are produced. Each pause in production has been argued to be due to the participant using search strategy in order to retrieve appropriate words (Gruenewald and Lockhead, 1980). Chertkow and Bub (1990) proposed a model of verbal fluency consisting of two components, ‘store’ and ‘search’. They found that Alzheimer’s patients, who tend to perform below average in verbal fluency tasks, suffered from inefficient search and/or damaged word storage. Meanwhile Troyer, Moscovitch and Wincour (1997) referred to storage as ‘cluster’ (the production of words from specific phonemic and semantic categories) and search as ‘switching’ (the process of efficiently shifting between strategies). As a result of proposals such as these, assessment of performance on the COWAT usually involves counting the number of clusters (both phonetic and semantic) and switches used, as well as counting the number of words produced.

COWAT assesses the ability to explore and retrieve appropriate words from long-term memory or the mental lexicon when a single letter is presented (Ruff et al., 1997). COWAT has also been reported to be a reliable measure of executive functioning, requiring functions such as: working memory, sustained attention, set-shifting and inhibition (Hedden and Yoon, 2006; Salthouse, Atkinson and Berish, 2003). As the ability to search for and retrieve words involves executive functioning, COWAT has become a popular neuropsychological task for assessing frontal lobe dysfunction.

The two most popular verbal fluency tasks are phonemic and semantic (generating words from a specific semantic category) fluency tasks (Ross, 2003). Research has shown that individuals with lesions to the left DLPFC and superior-medial suffer from poor performance on switching during phonemic fluency tasks, but have no deficits for semantic fluency tasks (Troyer et al. 1998). Troyer et al. (1997) argued that the process of switching (and thus switching scores) is related to frontal lobe functions such as set-shifting, cognitive flexibility and strategic searching. Clustering is associated with the temporal lobe functions such as semantic and lexical storage, as well as verbal memory. This may explain how lesions in different areas affect different phonemic and semantic verbal fluency tasks differently. There is support for this position from research on people with Huntingdon's disease (Rich et al., 1999), schizophrenia (Robert et al., 1998), Alzheimer's and Parkinson's (Troyer et al., 1998), as well as neuroimaging research (Drane et al., 2006).

A meta-analysis of phonemic fluency studies by Henry et al. (2004) revealed that individuals with frontal lobe lesions suffered from severe deficits in verbal fluency compared to healthy controls. This is supported by other lesion studies that have found that individuals with frontal lobe lesions produce fewer words during the COWAT than healthy controls (Baldo et al., 2001; Butler et al., 1993; Stuss et al., 1998). COWAT has also been used to compare verbal fluency between individuals with frontal and non-frontal lesions, with results revealing that individuals with frontal lesions produce fewer words than those with non-frontal lesions (Bornstein, 1986; Crockett et al., 1986; Milner, 1964).

Neuroimaging studies have also show that the frontal lobes are recruited during the COWAT. Warkentin and Passant (1997) reported increased activity in the left DLPFC, SMA and Broca's Area during the COWAT. Other neuroimaging studies have reported activity in the left DLPFC but not the right during the COWAT (Frith, 1995). The findings of these neuroimaging studies regarding the role of the left DLPFC in the COWAT are supported by lesions studies that have also noted the effect that left DLPFC lesion have on COWAT performance. A number of studies have found that deficits to phonemic fluency, including on the COWAT, are due to lesions of the left DLPFC (Baldo et al., 2001; Stuss et al., 1998; Troyer et al. 1998), with sufferers displaying deficits in comparison to patients with right DLPFC lesions.

5.3 Study 3

The aim of Study 3 is test the ability of fNIR to record haemodynamic changes in the DLPFC in order to validate the fNIR device. This will be achieved by using the three frontal tasks detailed above. This study will utilise computerised versions of the WCST, CBT task and COWAT, during which haemodynamic activity in the frontal lobes will be monitored using fNIR. This study also provides a means of evaluating the usage of fNIR in relation to frontal tasks. This is important as these tests are largely used to test clinical populations, meaning that the use of a new neuroimaging technique may help in the diagnosis and treatment of patients. So far there has only been limited research undertaken with fNIR of these frontal tasks. Hashimoto, Uruma and Abo (2008) used fNIR with the WCST and reported that individuals who had suffered traumatic brain injury had lower total haemoglobin levels in the PFC compared to control subjects. Shimodera et al. (2012) used fNIR to compare haemodynamic activity in the frontal lobes of schizophrenics and healthy controls during the COWAT. They found reduced haemodynamic activity in the frontal lobes of the schizophrenics, leading the authors to propose that fNIR is an effective measure of frontal lobe dysfunction. This is supported by research from Izekawa et al. (2009) who reported that schizophrenics displayed reduced PFC activity during a verbal fluency task when compared to healthy controls.

A computerised version of the WCST run using e-prime will be used as this provides a means with which to strictly control the timing and order of the task. The design will be modelled upon that used by Lie et al. (2006), except that this study will forego condition 'C' from their study, with only the 'A', 'B' and 'HLB' conditions being employed. This decision was taken in order to stop the testing procedure from lasting too long as the participants were also required to complete the WCST and the COWAT in a single testing session. Task C was omitted as task B provided a task with the informed switching element of task C, while the HLB contained perfectly matching reference cards that provided a similar level of challenge to the informed matching rule provided for each turn in task C. This means that the WCST task will have a difficulty hierarchy of A>B>HLB that is representative of the executive functions required for successful responses. Task A is a standard WCST, task B is a WCST in which the participant is informed of the sorting criteria every four turns and the HLB (the high level baseline) task is a simple matching task in which one of the stimulus cards matched one of the reference cards exactly on each dimension. It is hypothesised that there will be increased levels blood oxygenation for tasks A and B in comparison the HLB, with the highest overall levels found in task A as this task requires more complex cognitive functions. It is also predicted that performance on the WCST will be best for the HLB, followed by task B and then task A.

The CBT task will be implemented in the form of a computerised version run on e-prime. The design is based upon that used by Toepper et al. (2010). Three different Corsi tasks will be used: a standard CBT task, a BST and a baseline task in which the spatial location of the target block remains the same. Sequence lengths of 4, 5, 6, and 7 will be used for all of the tasks. In line with findings of Toepper et al. (2010) it is predicted that there will be increased levels of blood oxygenation for the CBT task and BST in comparison the baseline task, with the highest overall levels found in the BST. It is also predicted that performance on the CBT will be best for the baseline, followed by task the CBT and then the BST.

The COWAT will also be a computerised version run using e-prime as this method provided increased precision for the timings of the tasks. The verbal fluency task will be compared with a control condition in which participants will need to read pre-selected words aloud. Both tasks will last for 60 seconds. The hypothesis is that here will be increased levels of blood oxygenation for the verbal fluency task in comparison to the reading condition.

Given that previous laboratory based studies using fNIR have focused on reporting changes in blood oxygenation values, this study will also report changes in blood oxygenation values.

5.4 Methodology

5.4.1 Participants

A power analysis revealed that for a large effect size (F = 0.4), a total of 23 participants would be needed for the study.

27 participants (10 males and 17 females) took part in the study. All of the participants were Undergraduate students from the University of Nottingham. 20 of the participants were recruited through the school of Psychology's Research Participation scheme, a mandatory module component that provides the students with course credits. The other participants were recruited on an opportunity basis from other Undergraduate courses within the University. The average age of the participants was 19.44 years, with an age range of 18-21. All of the participants were right-handed, as assessed by the Edinburgh Questionnaire of Oldfield (1970). The participants had normal or corrected to normal vision, as well as no clinical memory deficits. None of the participants had any previous experience of any of the cognitive tests used in the study.

5.4.2 Apparatus and Stimuli

5.4.2.1 fNIR Device

The fNIR device (fNIR Devices, Ptotmac, MD, USA) used in this experiment belongs to the 'Continuous Wave' family of NIRS devices. Continuous wave fNIR uses light sources that emit light at a constant frequency and amplitude. The fNIR device consists of a 16-channel forehead sensor pad, an fNIR imaging control device and a computer with software for storing and analysing the fNIR data. The forehead sensor pad consists of 4 light sources (730nm/850nm dual wavelength LED) with an inter-optode distance of 25cm, as well as 10 light detectors (silicon photodiode with integrated trans-impedance preamp). The fNIR imaging control device sends power to the forehead sensor and then receives the data from the sensor, which it then sends to the computer for processing. The software for recording and analysing the fNIR data is the Cognitive Optical Brain Imaging (COBI) Studio (fNIR Devices, Ptotmac, MD, USA). The software utilises a modified Beer-Lambert Law in order to calculate changes in oxygenated and deoxygenated haemoglobin values relative to a pre-recorded baseline. Hb and HbO_2 values are calculated every 0.5s (2Hz sampling rate) relative to a pre-recorded baseline at each of the light detectors, with the light sources activating sequentially at both wavelengths every half-second. The device is capable of recording haemodynamic activity at cortical depths of 1-2cm.

5.4.2.2 *General Instructions*

A set of general instructions was used to explain the overall structure and purpose of the three tasks and can be found in Appendix 5.1.

5.4.2.3 *WCST*

For the WCST, the apparatus and stimuli included: a computer running e-prime, a computer keyboard and the visual stimuli used for the WCST. The order in which the participants completed the tasks was fully counterbalanced. Standardised instructions were provided that contained detailed instructions explaining what each of the WCST entailed and what the participants would be required to do (Appendix 5.2). The computer keyboard was modified by placing orange stickers over the 'X', 'C', 'B' and 'N' buttons, as these were the response buttons for the WCST (i.e. the far left button was for choosing the far left reference card). Participants were required to use their left hand for 'X' and 'C' and their right for 'B' and 'N'. The visual WCST stimuli were the same as those used by Lie et al. (2006). The stimuli consisted of 4 reference cards at the top of the screen that remained in the same location and order for every trial. The reference cards differed in terms of shape, colour and number of objects on the cards themselves. These objects were triangles, circles or stars. Below the reference cards was the stimulus card. During conditions A and B the stimulus card was a variant of the reference cards (via any combination of colour, shape or number). There were a total of 60 possible stimulus cards. For the HLB, the stimulus card was an exact replica of one of the reference cards. There were a total of 4 stimulus cards for the HLB. The order in which the stimulus cards were presented was randomised in each of the conditions. The stimulus dimension that changed (i.e. the colour, shape or number) was also randomised.

5.4.2.4 CBT

For the CBT task, the apparatus and stimuli included: a computer running E-prime, a keyboard and the visual stimuli for the CBT tasks. Standardised instructions were provided which contained detailed instructions explaining what each of the CBT entailed and what the participants would be required to do (Appendix 5.3). The task stimuli followed the design originally used by Corsi (1972), with the experimental procedure matching that of Toepper et al. (2010). This consisted of a black (RGB 0 0 0) Corsiboard, with claret red (RGB 163 163 163) blocks placed upon the Corsiboard. Target blocks appeared in natural red (RGB 255 0 0), while BST distractor blocks appeared in yellow (RGB 255 255 0). Response blocks were coloured grey (RGB 163 163 163) with yellow letters 'L' and 'R' displayed upon them. Each of the blocks measured 3x3cm. A fixation cross was also used. This was placed at the centre of the screen, in black on a white (RGB 255 255 255) background. The task types, sequences, and spatial locations of the target blocks and response option blocks was fully counterbalanced and randomised. The tasks were run using e-prime. A button response pad was connected to the computer and used for making responses to the response blocks. The left button corresponded to the 'L' block while the right button corresponded to the 'R' block.

5.4.2.5 COWAT

The COWAT used: a computer running E-prime, a voice recorder and a set list of words for the control condition. Standardised instructions were provided which contained detailed instructions explaining what each of the COWAT entailed and what the participants would be required to do. These instructions were based upon those that are used in standardised versions of the COWAT (Spreen and Strauss, 1998) and can be seen in Appendix 5.4. The voice recorder was used so that the participant's responses could be analysed using the method of Troyer and Moscovitch (1997). Each of the four letters used ('F', 'A', 'R' and 'S') had a set list of words that were used in the reading condition (Condition A). The words were presented in a random order on the computer monitor, with a fixation cross appearing in between the presentation of the words. The words were written in black letters on a white background. These words were collected as part of a pilot study conducted on 20 participants; in order to collect common words produced beginning with the letters used for this COWAT. A total of 48 words were used for Condition A, 12 words for each letter. Condition A was run using e-prime. The list of words used can be seen in Condition A can be found in Appendix 5.5.

5.4.3 Design

5.4.3.1 General

The overall design of the study (including the WCST, CBT task and COWAT) was a within-subjects design, with all participants taking each of the cognitive tasks.

5.4.3.2 WCST

The independent variable in the WCST was what type of WCST the participant was completing. This independent variable had 3 levels: standard version of the original WCST (Condition A), a version in which the participants were informed of the sorting criteria every 4 turns (Condition B), or a simple matching version in which one of the stimulus cards matched one of the reference cards exactly on each dimension (HLB). Conditions A and B were completed as part of separate runs, each alternating with blocks of the HLB. Each of the runs, A-HLB and B-HLB, consisted of 4 HLB blocks that alternated with experimental condition blocks (i.e. A or B). HLB baseline blocks appeared at the beginning and end of each run. Each run consisted of 7 blocks, while each block consisted of 12 trials, meaning that each run contained of 84 trials (36 trials of the experimental condition and 48 HLB). Each individual trial lasted for 4.75s, resulting in a block length of 57s (12x4.75s). As a result each run lasted 6 minutes 39s (as there are 7 blocks in each run). The experiment was designed so that A-HLB was always followed by B-HLB, as B-HLB may have enabled the participants to deduct the sorting criteria of A-HLB, for which they were meant to be naïve of at the start of the experiment.

The dependent variables for the WCST were the haemodynamic measures recorded using fNIR and the behavioural data calculated from the participant's performance on the WCST. The haemodynamic measures consist of the changes in Hb, HbO_2 and blood oxygenation during the three WCST tasks. These values are calculated in relation to a baseline recorded before the start of the WCST task. The behavioural measures consist of: cognitive costs, accuracy, preservative errors, set loss errors and other errors.

While task B and the HLB are different than the standardised WCST, task A is fundamentally the same in nature to the standard WCST as it involves uninstructed set shifting. The only difference between task A and the standard WCST is a reduction in the overall number of trials and dimensional sets used, which is due to this study also using task B and the HLB from Lie et al. (2006).

5.4.3.3 CBT

The independent variables for the CBT were the task type and sequence length of the task. Task type had 3 levels: a standard CBT task, a BST or a baseline block-tapping task. The standard CBT task required the participants to identify the sequence in which spatially separated blocks had been illuminated. The BST is the same as the standard CBT task except this time a distractor block is illuminated at the same time as each target block. In the baseline task the illuminated target blocks remained the same in each trial. Sequence length had 4 levels with 3, 4, 5, or 6 blocks.

For each of the tasks the trials were divided into a 'encoding' and 'response' phase. The encoding phase consisted of a blank Corsiboard (presented for 1.5s), followed by the illumination of the target blocks (lasting for 0.5s each), followed by again by a blank Corsiboard (presented for 2s). After this the response phase took place. The response phase consisted of the same Corsiboard as was used in the encoding phase, with two of the block illuminated, one with the letter 'L' on it and the other with the letter 'R' on it. The participant had to pick which one of the blocks had been a part of the sequence in the encoding phase. The response phase consisted of 3, 4, 5, or 6 choices depending on the sequence length. The response phase had a set time limit of 6s, 8s, 10s or 12s for sequence lengths of 3, 4, 5 and 6 respectively. This meant that the participants had a total of 2s to respond for each block choice. If they did not respond within 2s then the response option moved to the next block in the sequence. After the response phase had ended a fixation cross was displayed before the next encoding phase began (for the next sequence length).

Each sequence length was conducted for each task type in one block. In total there were 12 trials per block. As a result each participant completed 48 trials (4 sequence lengths x 3 task types x 4 blocks). Total trial durations were 12s, 15s, 18s and 21s for sequence lengths of 3, 4, 5, and 6 respectively. The total task duration was 16.5 minutes. This includes the time that the participants had for a practice block to familiarise themselves with the CBT tasks. The practice block consisted of 12 trials, 1 for each sequence length across each of the three task types. The practice trial lasted for a total of 3.3 minutes.

The dependent variables for the CBT task were the haemodynamic measures recorded by fNIR and the behavioural measures from the CBT task. The haemodynamic measures consist of the changes in Hb, HbO_2 and blood oxygenation during the three WCST tasks. These values are calculated in relation to a baseline recorded before the start of the CBT task. The behavioural measures were the percentage of correct responses and the reaction time for correct responses.

Besides the use of the BST and the baseline condition, the CBT used in this study differs from the design of the standardised version in a few ways. In the standardised version (Corsi, 1972) the participant reproduces the sequence by touching - or in the case of computerised versions, clicking - the Corsi blocks. The participant also responds by reproducing the entire sequence without any set response periods for each block in the sequence. In this study responses were made using alternate choice button response, with two blocks illuminated and the participants required to pick either the ‘L’ or the ‘R’ block. There were also set response periods for each of the blocks, meaning that the participants could not go from one block to the next at their own pace. The reason for this design choice was so that accurate response times could be recorded. Response times are not usually recorded for the CBT (accuracy is the usual measure). In this study response times were used as a behavioural measure of the executive functions of the BST. In order to make sure that the response method was the same for all three conditions, this response method was also used for the CBT and baseline task. This is the exact same design and rationale that was used with success by Toepper et al. (2010), who designed the methodology with neuroimaging in mind. The use of strict timings and responses allows for neuroimaging data, in this case fNIR, to be synchronised with the events taking place in the task. In particular it allows for neuroimaging data to be synchronised with the encoding and response phases.

5.4.3.4 COWAT

The independent variable for the COWAT was the task type. This had 2 levels, whether the participant had to read aloud words provided to them (control condition, Condition A) or had to verbally generate their own words using a letter provided to them (Condition B). During Condition A, participants were presented with 12 words that each began with the same letter over 60s (each word was presented for 3s, with a fixation cross presented for 2s between each word). For Condition B, participants were provided with a letter and then given 60s to generate as many words as possible. Participants completed 4 runs in total, two for Condition A and two for Condition B. The order in which the participants completed the runs was counterbalanced so that the runs were either in the order ABBA or BAAB (with half of the participants completing the ABBA run and the other half completing the BAAB run). The ABBA run used the letters ‘F’ and ‘S’ for Condition A and the letter ‘A’ and ‘R’ for Condition B. This was vice versa for the BAAB run order.

The dependent variables for the COWAT were the haemodynamic measures recorded by fNIR and the behavioural measures from the COWAT. The haemodynamic measures consist of the changes in Hb, HbO_2 and blood oxygenation during the three WCST tasks. These values are calculated in relation to a baseline recorded before the start of the COWAT. The behavioural measures consisted of: number of correct words, phonetic categories, semantic categories, mean cluster size and switches. These measures were analysed according to the method proposed by Troyer and Moscovitch (1997).

The COWAT design used in this study was similar to the standardised versions of the test that are widely used today. When compared to the original COWAT (Benton and Hamsher, 1976; Benton, Hamsher and Sivan, 1994), the biggest change was the use of a reading task to accompany the verbal fluency task. This was necessitated by the fact that this study was primarily interested in validating the fNIR device, which required a baseline condition (the reading condition) for comparison with the verbal fluency task. A further change involved only using two letters for the verbal fluency task, with a further two letters used for the reading task. The decision to use four letters across both tasks was taken in order to stop the testing session from getting too long and in order to allow for a counterbalanced ABBA/BAAB design to be used. The use of four letters also allowed for the overall design of the study (both word production and reading) to remain similar to common versions of verbal fluency tests, which at most use four letters (e.g. the Test of Verbal Conceptualisation and Fluency from Reynolds and Horton, 2006). This study also differed from the original COWAT in terms of the letters used. Instead of using the letters 'C', 'F' and 'L', this study used the letters 'F', 'A' and 'S' as these are now the most widely used letters for the COWAT (Spreen and Strauss, 1998). The letter 'R' was also used in addition to these letters in order to allow for there to be an equal number of letters for the word production and reading tasks. 'R' was chosen, as it is a letter relatively commonly used in verbal fluency tasks (Spreen and Strauss, 1998).

5.4.4 Procedure

5.4.4.1 General

Before starting any of the tasks, the experimenter provided a general introduction by explaining what the purpose of the study was as well as explaining the fNIR device to the participants. The participants were also given a set of general instructions. The order in which the participants completed the WCST, CBT task and COWAT was fully counterbalanced. The participants sat down in front of the computer monitor and were then given the chance to ask any questions regarding the experiment. Before the participant began their first task the fNIR sensor was secured on their forehead (making sure that there was no hair in the way) using Velcro straps and a cloth (to help keep out external light). An fNIR reading was then taken to make sure that a clear signal was being received. The signal was checked by examining the raw fNIR signals displayed by the fNIR control software. The software displays the raw signals for the light at 730nm and 850nm. These signals must be within a range of 1000 to 4000mV in order to be regarded as valid and reliable signals. If this was not the case, the sensor pad was re-placed on the forehead as a bad signal often results from the sensor pad not being firmly secured to the forehead or hair obstructing the sensors. If the signal was still not at an appropriate level, the gain settings of the fNIR device were adjusted until the signal was within the acceptable boundaries.

Once a good signal was achieved it was possible to take baseline readings before each task began. Taking a baseline consisted of having the participants stare at a blank computer monitor for 10s while the fNIR device carried out baseline calculations. Recording changes in haemodynamic activity during the tasks involved setting the device to record at the beginning of a run or condition and stopping the device from recording at the end of a that run or condition. The number of baselines and recordings taken varied between the WCST, CBT task and the COWAT. More detailed explanations can be found in the respective procedure sections.

5.4.3.2 *WCST*

Before starting the WCST the participants were provided with the standardised instructions that explained each of the tasks and what they would be required to do. The participants were instructed that they needed to match one of the cards on the bottom row to one of the cards on the top row. The participants then had the opportunity to ask the experimenter any questions they had regarding the tasks. The WCST was split into two parts, with participants first completing the A-HLB condition, and then completing the B-HLB condition. The participants were instructed to use their left index and middle fingers to press the left buttons when they wanted to pick one of the reference cards on the left. They were instructed to do the same with their right hands to pick one of the reference cards on the right. Before the start each A or HLB block, the participants were informed by onscreen instructions of what they were required to do. For task A the participants had to determine the sorting criteria by themselves. For the HLB task the stimulus card would always perfectly match one of the reference cards. After each response the participant was given the feedback ‘Correct’ or ‘Incorrect’. For task A the selection criteria would change after four consecutive correct responses. There were 4 blocks of task A and the HLB each. Each block of A or HLB consisted of 12 trials, after which the task switched. This meant that there was a maximum of 3 criterion changes for task A in each block.

After the A-HLB condition had been completed, the participant moved onto the B-HLB condition. The HLB task remained the same as in the A-HLB condition. The participants were again informed of their task before the start of each block. For task B the participants were informed that they would be told the sorting criteria prior to the 4 trials that criteria applied to. A message detailing the selection criteria would appear on the screen saying ‘SHAPE’, ‘COLOUR’ or ‘NUMBER’. After each response the participant was given the feedback ‘Correct’ or ‘Incorrect’. After 4 trials were completed the selection criterion would change. Each block of B comprised 12 trials, meaning that there were 3 selection criteria in total. After each response (in all three conditions) an onscreen message appeared telling the participant if the card they had chosen was ‘CORRECT’ or ‘INCORRECT’.

fNIR baselines were taken twice, once before the A-HLB condition and again before the B-HLB condition. Changes in haemodynamic activity were recorded throughout both conditions, resulting in two sets of fNIR data. During the baseline recordings a message on the computer monitor informed the participants that an fNIR baseline was being taken.

5.4.3.3 CBT

Before starting the CBT task, the participants read the standardised instructions that explained the three different CBT tasks as well as what they were required to do. The participants then had the chance to ask the experimenter any questions they had regarding the tasks. The experimenter instructed the participants that their task would be to remember the order in which a set of blocks illuminated and then respond on the basis of that order. The participants were first given the opportunity to practice the tasks by completing a practice task. The participants were instructed to use the 'm' key on the keyboard to pick the block with 'L' on it, and to use the 'z' key on the keyboard to pick the block with 'R' on it. The practice task involved trials covering each of the task types for each of the sequence lengths (12 trials in total). After completing the practice trials the participants moved onto the experimental portion. The CBT task involved 4 blocks, each consisting of 12 trials. The 12 trials covered the three CBT tasks (the standard CBT task, BST and the baseline) for the four sequence lengths (3, 4, 5 and 6 blocks). The response the participants needed to make was the same for each task type, with the participants needing to decide whether the 'L' block or the 'R' block was a part of the sequence in the encoding phase. The task types, sequences, and spatial locations of the target blocks and response option blocks was fully counterbalanced and randomised.

fNIR baselines were taken before each of the four blocks, with changes in haemodynamic activity being recorded over all twelve trials of each block. This resulted in four sets of fNIR data, one for each block. During the baseline recordings a message on the computer monitor informed the participants that an fNIR baseline was being taken.

5.4.3.4 COWAT

Before starting the experiment the participants read the standardised instructions that explained the COWAT and what they would be required to do. The participants were also given the opportunity to ask the experimenter any questions that they had. The COWAT was first set up according to whether the participant was completing the ABBA or BAAB order. A message on the computer monitor first informed the participants that an fNIR baseline was being taken. After the baseline was taken an information screen appeared explaining the task (there were two sets of instructions, one for Condition A and another for Condition B). For Condition A the participants were informed that they needed to verbally produce as many words as possible using a single letter in 1 minute, while avoiding repetition, homophones and multiple words with the same stem. For Condition B the participants were informed that they needed to read aloud words that were subsequently displayed on the computer monitor. Each participant completed Conditions A and B twice. Separate fNIR recording were made across all four runs. This involved taking baselines before the start of each of the runs. While the baselines were being recorded, the participants were instructed to stare at the computer monitor (which informed them an fNIR baseline was being taken. Separate recordings of haemodynamic changes were taken for each of the runs, with the device being set to record at the same time that the task began. Upon the completion of that particular run, the fNIR device was stopped from recording. This resulted in four sets of baselines and recordings.

5.5 Results

5.5.1 WCST

5.5.1.1 Behavioural Data

The scoring of the behavioural data for the WCST was based upon those used by Heaton et al. (1993) and Lie et al. (2006).

5.5.1.2 Cognitive Costs

Cognitive costs were defined as the mean correct reaction time (in ms) for Condition A and B minus the mean correct reaction time for the HLB in each of the respective conditions. (HLB reaction times were treated separately for Conditions A and B after an independent *t*-test revealed them to be significantly different $t(26)= 6.79, p<.001$.

Cognitive costs were significantly different between task A and B $t(26)= 5.76, p<.001$, with reaction times higher for task A (582ms) than task B (295ms).

5.5.1.3 Accuracy Measures

Accuracy measures were defined as the total number of errors the participants made in Condition A and Condition B. Accuracy measures were found to be significantly different between task A and B $t(26)= 17.84$, $p<.001$, with more errors made during task A(M=5.04) than task B (M=0.69).

5.5.1.4 Preservative Errors

Preservative errors were errors in Condition A in which the participants responded on the basis of a previously 'correct' criteria which no longer applied despite having been informed that they had made an incorrect selection. The mean number of preservative errors was 1.95 (range 0-5).

5.5.1.5 Set Loss Errors

These were errors in Condition A determined by the number of errors occurring after three consecutively correct responses which are though to measure the understanding and deduction of the correct sorting criteria by the participant. The mean number of set loss errors was 0.12 (range 0-1).

5.5.1.6 Other Errors

This includes any errors that are not consistent with any of the above measures. The mean number of other errors was 2.96 (range 1-9).

5.5.1.7 Correct Trials and Reaction Times

There was a significant difference between the number of correct dimensional sets $t(26)= 11.71$, $p<.001$, with subjects achieving an average of 1.1 complete sets for task A compared to 2.2 for task B. None of the participants were able to complete all three-dimensional sets in task A, whereas 14.8% achieved this for task B.

When making correct responses during trials, participants were significantly faster in doing so during task B than in task A $t(26) = 6.79$, $p < .001$. There was an improvement in reaction times for participants as they progressed through the three experimental blocks however.

The mean number of correct trials for task A was significantly lower compared to HLB(A) $t(26) = 18.04$, $p < .001$. Reaction times for correct responses were also slower during task A compared to HLB(A) $t(26) = 12.28$, $p < .001$.

The mean number of correct trials for task B was significantly lower compared to HLB(B) $t(26) = 4.38$, $p < .01$. Reaction times for correct responses were also slower during task B compared to HLB(B) $t(26) = 13.38$, $p < .01$.

Table 5.1 provides a full summary of the correct trials and response times.

Table 5.1. Summary of correct trials and response times for the WCST.

Task	Block Number	MeanCorrect Dimensions	MeanCorrect ReactionTime (ms)	MeanCorrect Trials
A	1	1.07	1510.92	6.85
	2	1.11	1353.90	6.78
	3	1.11	1292.40	7.26
B	1	2.33	976.58	11.30
	2	2.19	1015.37	11.41
	3	2.07	1015.89	11.22
HLB(A)	1	2.7	943.39	11.63
	2	2.81	774.43	11.81
	3	2.85	775.80	11.81
	4	2.67	740.57	11.67
HLB(B)	1	2.85	538.33	11.85
	2	2.7	513	11.70
	3	2.85	518.33	11.78
	4	2.81	502.33	11.78

5.5.2 CBT

5.5.2.1 Behavioural Data

The scoring of the behavioural data for the CBT tasks was based upon that used by Toepper et al. (2010).

5.5.2.2 Proportion Correct

The proportion of correct responses was averaged across each condition and sequence length. A 3x4 (task x sequence length) repeated-measures ANOVA revealed a significant main effect of condition F (2,52) = 34.136, MSE = 66.831, p <.001, ${}_p\eta^2$ =.57. Planned Helmert contrasts revealed that there were significant differences between the baseline task and the CBT and BST F (1,26) = 42.488, MSE = 38.595, p <.001, ${}_p\eta^2$ =.62, with more correct responses for the baseline, and that the CBT and BST were significantly different from each other F (1,26) = 6.175, MSE = 15.371, p <.05, ${}_p\eta^2$ =.19, with more correct responses for the CBT. A significant main effect of sequence length was also found F (3,78) = 19.846, MSE = 51.361, p <.001, ${}_p\eta^2$ =.43, with a planned Helmert contrast revealing that all of the sequence lengths were significantly different from each other F (1,26) = 30.408, MSE = 31.779, p <.001, ${}_p\eta^2$ =.54, with the number of correct responses decreasing with increasing sequence length. A significant task by sequence length interaction was found F (6, 156) = 6.676, MSE =44.201, p <.001, ${}_p\eta^2$ =.21 and can bee seen in Figure 5.1. Planned Helmert contrasts revealed that compared to baseline, the proportion of correct responses for CBT and BST were significantly different at sequence lengths 5 F (1,26) = 20.833, MSE = 46.945, p <.001, ${}_p\eta^2$ =.44 and 6 F (1,26) = 21.504, MSE = 87.195, p <.001, ${}_p\eta^2$ =.45, with fewer correct responses at these sequence lengths. A planned Helmert contrast also revealed that the CBT and BST were significantly different at sequence length 5 F (1,26) = 7.3, MSE = 108.520, p <.05, ${}_p\eta^2$ =.22, with fewer correct responses being made during the BST.

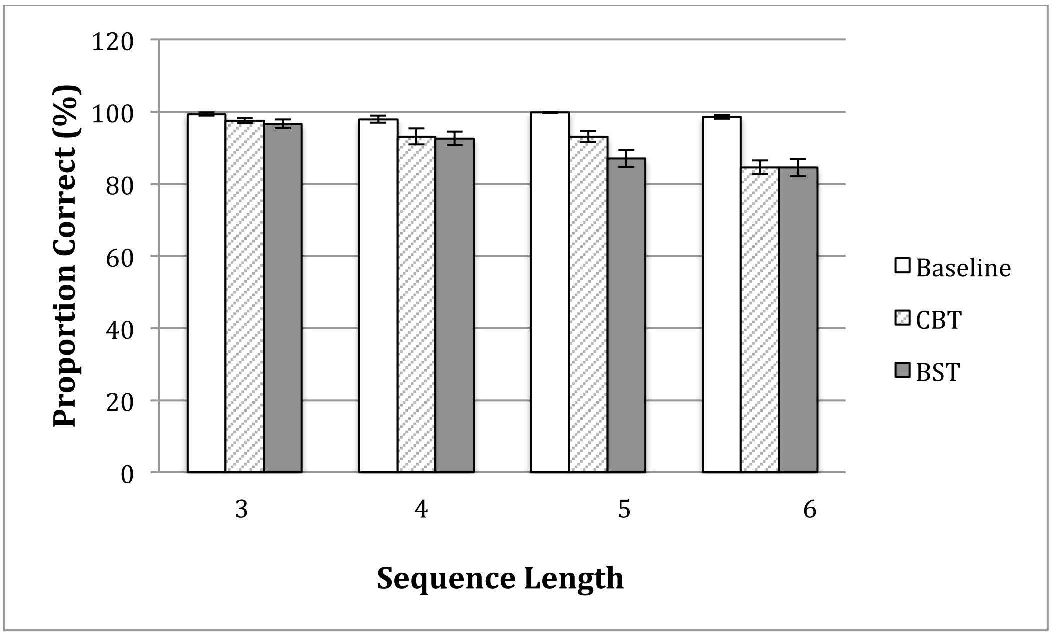


Figure 5.1. Significant interaction between task and sequence length on the proportion of correct responses. Error bars represent +/- 1 S.E.

5.5.2.3 Correct Reaction Time

Reaction time (in ms) for correct responses was averaged across each condition and sequence length. A 3x4 (task x sequence length) repeated-measures ANOVA revealed a significant main effect of task F (2,52) = 19.206, p <.001, MSE =, $_p\eta^2$ =.42, with a Helmert contrast revealing that the baseline task was significantly different from CBT and BST F (1,26) = 37.312, p <.001, $_p\eta^2$ =.59, with shorter reaction times for the baseline. The CBT and BST were not significantly different however [F (1,26) = 3.498, p <.073, $_p\eta^2$ =.03]. A significant main effect of sequence length was found F (3,78) = 31.484, MSE =, p <.001, $_p\eta^2$ =.55, with Helmert contrast revealing that all sequence lengths were significantly different from each other F (1,26) = 47.58, p <.001, $_p\eta^2$ =.64, with reaction times increasing with sequence length. A significant task by sequence length interaction was found F (6,156) = 7.026, MSE =, p <.001, $_p\eta^2$ =.21 and can be seen in Figure 5.2. A Helmert contrast revealed that compared to the baseline, CBT and BST were significantly different at sequences lengths of 5 F (1,26) = 28.467, p <.001, $_p\eta^2$ =.48 and 6 F (1,26) = 28.254, p <.001, $_p\eta^2$ =.52, with the reaction times being significantly slower for the CBT and BST. There was also a significant difference between CBT and BST at sequence lengths 3 and 4 $F\,(1,26) = 4.407$, $p < .05$, ${}_{p}\eta^2 = .15$, with reaction times significantly slower for the BST.

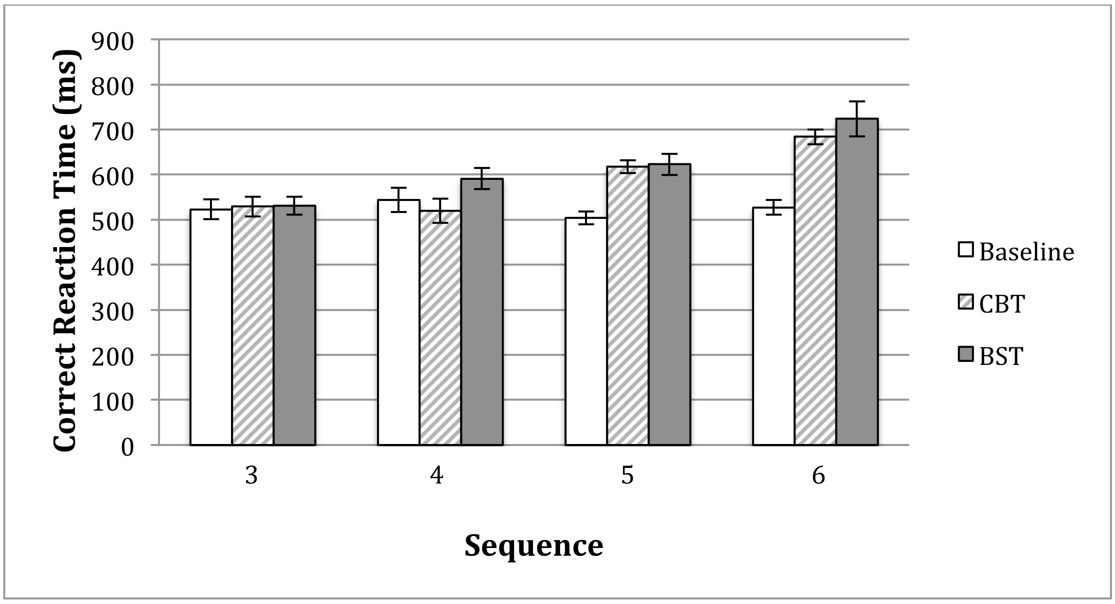


Figure 5.2. Significant interaction between task and sequence length on the correct reaction time. Error bars represent +/- 1 S.E.

5.5.3 COWAT

5.5.3.1 Behavioural Data

The scoring of the behavioural data was based upon the scoring rules proposed by Troyer and Moscovitch (1997).

5.5.3.2 Number of Correct Words

The total number of words produced in Condition A, excluding any incorrect words and repetitions. The number of correct words produced, averaged across letters, was 26.63 (range 9-57).

5.5.3.3 Phonetic Categories/Clusters

Words were assigned to this category if there were a collection of two or more words that demonstrated phonetic similarities (e.g. starting or ending in the same letter, rhyming words). A difference between this study and that of Troyer and Moscovitch (1997) was that the participants were instructed to avoid using homophones; with homophones being regarded as incorrect words. The mean number of phonetic categories produced was 4.19 (range 1-15).

5.5.3.4 Semantic Categories/Clusters

Words were assigned to this category if a collection of two words or more displayed a clear sematic correspondence. The mean number of semantic categories produced was 1.89 (range 0-5). The difference between the number of phonetic and semantic categories was significant $t(26)= 4.07$, $p<.001$.

5.5.3.5 Mean Cluster Size

To qualify as a phonetic or semantic cluster there needed to be at least two associated words. The size of an individual cluster was determined by the number of words succeeding the first word in the cluster (e.g. a cluster of three would have a cluster size of two). Cluster size was averaged across categories for each individual and then across all of the participants. The mean cluster size was 1.39 (range 1-2.29).

5.5.3.6 Mean Switches

Switches were defined as switches between any phonetic or semantic clusters, as well as between individual words that did not belong to any clusters. This included any incorrect words and repetitions. The number of switches was averaged across the participants. The mean number of switches was 16.93 (range 6-34).

A summary of the COWAT behavioural data can bee seen in Table 5.2.

Table 5.2. Summary of behavioural data from the COWAT.

	Correct words	Phonetic Groups	Semantic Groups	Cluster Size	Switches
Mean	26.62	4.19	1.89	1.39	16.92
S.D.	10.35	2.87	1.48	0.037	7.54

5.5.4 fNIR

The analysis of the fNIR data involved analysing the changes in blood oxygenation values throughout the different tasks. Blood oxygenation was calculated by subtracting the Hb values from the HbO_2 values at each recording time point (i.e. every half second). For the fNIR analysis, the 16 channels (voxels) were divided into four regions (each comprising four voxels) in an attempt to see of there were any differences in haemodynamic activity as a function of laterality or superiority. The four regions were: superior left, inferior left, superior right and inferior right.

5.5.4.1 WCST

Before the analysis began, two of the participants were omitted from the WCST analysis due to missing data. The HLB tasks for tasks A and B were treated separately in the analysis. The HLB will subsequently be referred to as HLB(A) and HLB(B). All of the tasks were analysed using a number of different 2x2x2 (task x superiority x laterality) repeated-measures ANOVAs.

A 2x2x2 repeated-measures ANOVA for task A and HLB(A) revealed a significant main effect of task F (1,24) = 4.34, p <.05, MSE = 4.571 $_p\eta^2$ =.15, with significantly higher blood oxygenation levels during task A than HLB(A). There was also a significant main effect of superiority F (1,24) = 8.50, MSE =4.569, p <.01, $_p\eta^2$ =.26 with higher levels of blood oxygenation in the lower voxels. However there was no significant main effect of lateralisation [F (1,24) = 0.442, MSE =0.760, p =.51, $_p\eta^2$ =.01]. A significant interaction between task and superiority was found F (1,24) = 6.99, MSE =0.211, p <.05, $_p\eta^2$ =.23, with task A producing higher levels of blood oxygenation in the inferior voxels compared to HLB(A). This can be seen in Figure 5.3.

A 2x2x2 repeated-measures ANOVA for task B and HLB(B) revealed a significant main effect of lateralisation F (1,24) = 4.54, p <.05, MSE = 1.91, $_p\eta^2$ =.16, with significantly higher oxygenation in the voxels on the left. There were no significant main effects of task [F (1,24) = 0.320, MSE =0.136, p =.57, $_p\eta^2$ =.04] or superiority [F (1,24) = 2.45, MSE = 1.949, p =.13, $_p\eta^2$ =.09] and no significant interactions.

A 2x2x2 repeated-measures ANOVA comparing task HLB(A) and HLB(B) revealed a significant main effect of superiority F (1,24) = 5.11, p <.05, MSE =2.329, $_p\eta^2$ =.17, with significantly higher levels of blood oxygenation in the inferior voxels. There were no significant main effects of task [F (1,24) = 0.000, MSE =0.003, p =.98, $_p\eta^2$ =.000] or laterality [F (1,24) = 1.268, MSE =0.854, p =.27, $_p\eta^2$ =.05] and no significant interactions.

A 2x2x2 repeated-measures ANOVA for task A and task B revealed a significant main effect superiority F (1,24) = 7.87, MSE = 4.03, p <.001, $_p\eta^2$ =.25, with significantly higher oxygenation in inferior voxels. Although not significant, [F (1,24) = 3.227, p =.09, $_p\eta^2$ =.1], an interaction between superiority and lateralisation trended towards higher levels of oxygenation in the left inferior voxels. No significant difference in task was found between tasks A and B [F (1,24) = 0.082, MSE = 1.806, p =.77, $_p\eta^2$ =.003]. No significant interactions were found either.

The mean level of oxygenation for each participant in tasks A and B was correlated with the mean cognitive costs and the accuracy measures. However no significant correlations were found for any of the measures.

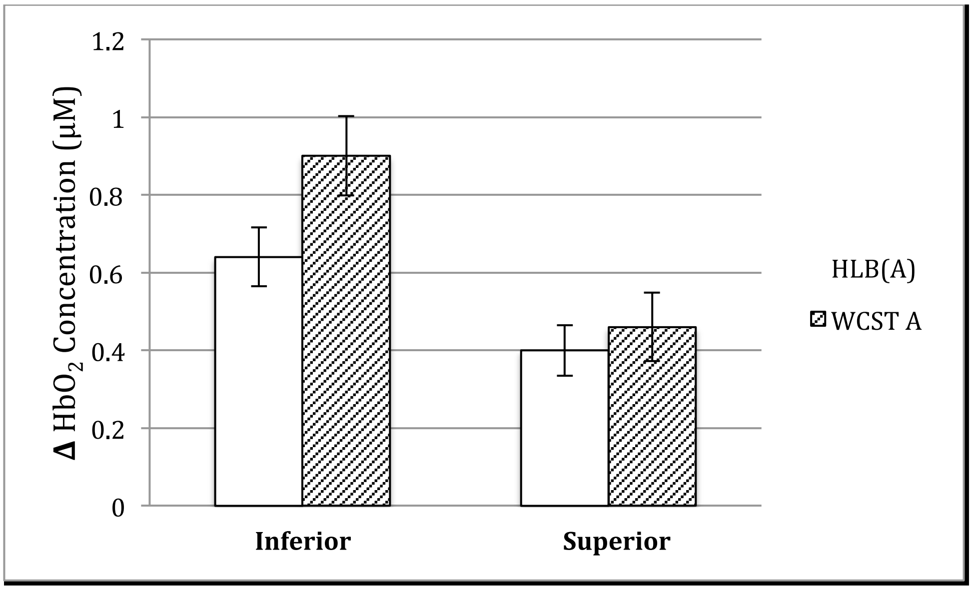


Figure 5.3. Interaction between change in blood oxygenation and voxel region for the HLB(A) and WCST A. Error bars display +/- 1 standard error.

5.5.4.2 CBT

The fNIR data was analysed using 3x4x2x2 (task x sequence length x superiority x laterality) repeated-measures ANOVA on the encoding and response phases. No significant main effects or interactions were found. Following the suggestions of Toepper et al. (2010) a further analysis was carried out using a reduced time window for the encoding phase only, however this still failed to produce any significant main effects or interactions (possibly due to increased variance and noise in the data due to decreasing the time window). There was a trend towards increased oxygenation during the BST but this was not significant [F (2,52) = 0.770, p =.51, $_{p}\eta^{2}$ =.02]. A summary of the changes in oxygenation during the Corsi tasks can be found in Table 5.3. These values make it clear that there was a very large amount of variability in the data in the three tasks.

Table 5.3. Changes in blood oxygenation for the three Corsi tasks (measured in micro mols). Values in parentheses are standard deviations.

	3 Blocks	4 Blocks	5 Blocks	6 Blocks
Baseline	0.04 (0.6)	0.07 (0.5)	0.1 (0.5)	0.1 (0.55)
CBT	0.075 (0.55)	0.1 (0.46)	0.1 (0.55)	0.08 (0.6)
BST	0.15 (0.5)	0.13 (0.5)	0.15 (0.38)	0.12 (0.55)

5.5.4.3 COWAT

The fNIR data was analysed using a 2x2x2 (task x superiority x laterality) repeated-measures ANOVA. The results revealed a significant main effect of task F (1,26) = 18.31, MSE =45.4, p <.001, $_{p}\eta^{2}$ =.41, with increased oxygenation levels during the generation task compared to the reading task. The results also revealed a significant main effect of superiority F (1,26) = 4.79, MSE = 1.036, p <.05, $_{p}\eta^{2}$ =.15, with inferior voxels displaying increased oxygenation levels that superior voxels. No effect of lateralisation was found [F (1,26) = 0.115, MSE = 0.019, p=.73, $_{p}\eta^{2}$ =.0.003]. No significant interactions were found either. Figure 5.4 details the mean change in blood oxygenation across the voxel regions for the generation and reading tasks.

A bivariate correlation revealed that there was a significant correlation in oxygenation levels between the voxel regions and mean cluster size: superior left $r(25) = 0.43$, $p<.05$, inferior left $r(25) = 0.41$, $p<.05$, superior right $r(25) = 0.44$, $p<.05$ and inferior right $r(25) = 0.41$, $p<.05$. No significant correlations were found between voxel groups and any of the other behavioural measures for the COWAT.

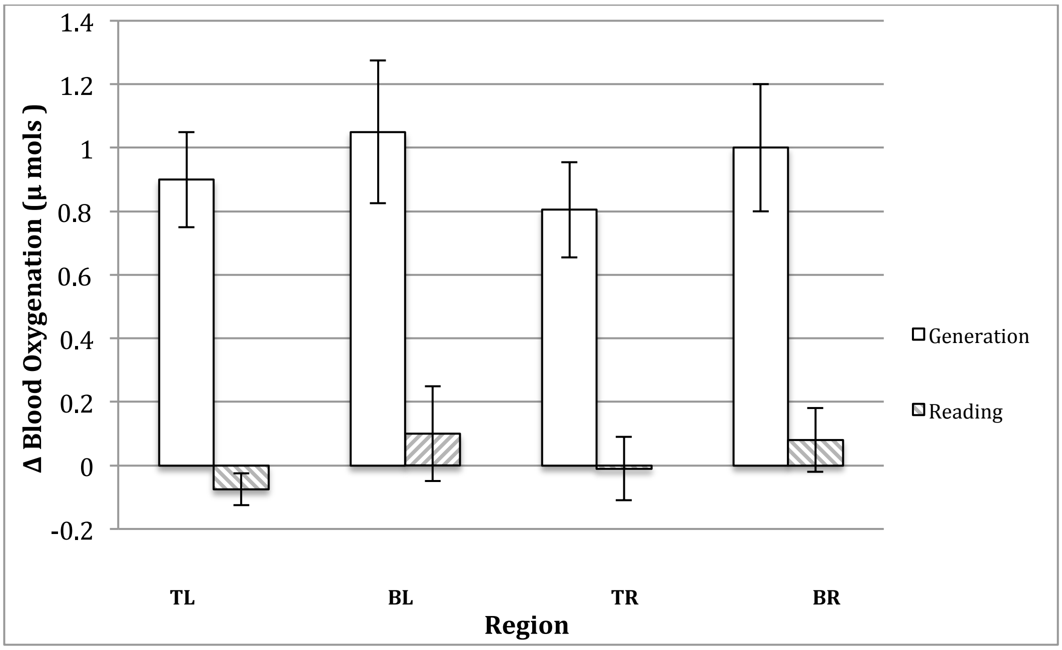


Figure 5.4. Mean change in blood oxygenation for the generation and reading tasks across the voxel regions. Error bars represent +/- 1 S.E.

Changes in blood oxygenation throughout the duration of the word generation and reading tasks were averaged across all participants and can bee seen in Figure 5.5. This demonstrates that blood oxygenation decreased throughout the duration of the reading task while it increased throughout the duration of the word generation task.

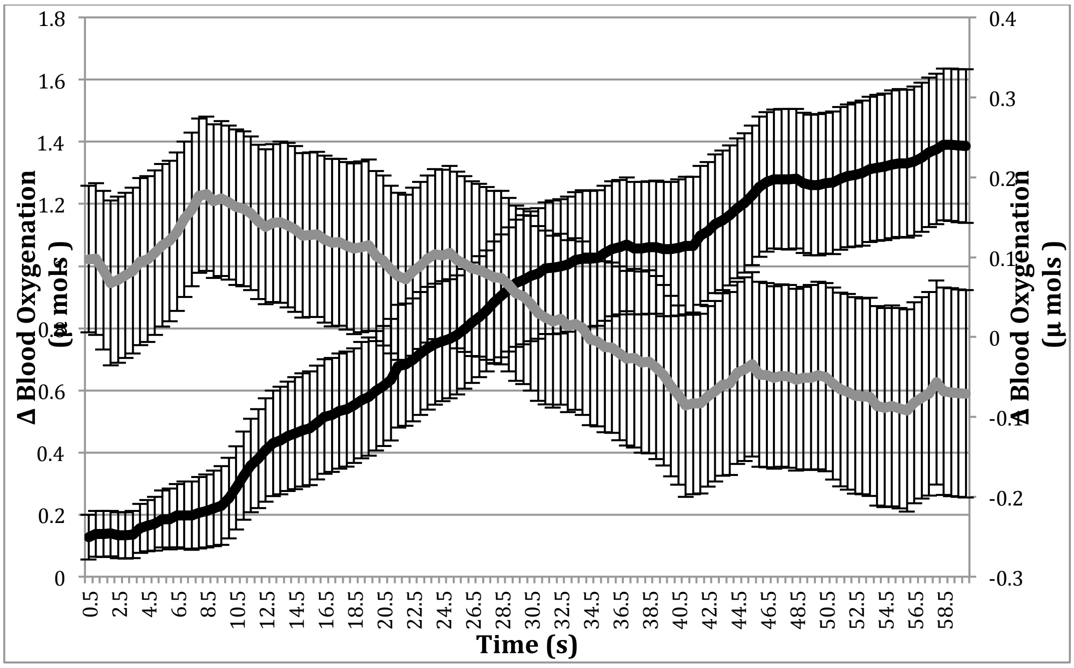


Figure 5.5. Mean changes in blood oxygenation for all participants during the word generation and reading tasks for the COWAT. Error bars display +/- 1 standard error. The black line represents the generation condition and the grey line represents the reading condition.

5.6 Discussion

5.6.1 WCST

The behavioural results of the WCST demonstrate that the design of the task was effective, with the behavioural results falling in line with the difficulty hierarchy (A>B>HLB). Task A was associated with higher cognitive costs, more errors, fewer correct sets and slower reaction times in comparison to task B. Meanwhile the HLB produced more correct sets and faster reaction times than tasks A and B. This shows that as expected the HLB was the easiest task of all. This demonstrates that the participants found task A more significantly more difficult than task B, as during task A they had to deduce the sorting rule for themselves whereas in task B the sorting criterion was provided to them for every set of four trials. This means that whereas task A required complex executive functioning in order to select the correct sorting criteria, task B simply required the participants to maintain the sorting criterion provided to them in working memory. The executive functioning required by the different tasks is based upon the levels of cognitive switching required: uninstructed (task A), instructed (task B) and no set-shifting (HLB). Cognitive switching entails a greater level of cognitive demand, which would explain the effect on task performance.

The slower reaction times for task A were indicative of the participants needing extra time to deduce the sorting criteria through the use of executive functions. The greater difficulty of task A also manifested itself in the increased number of errors. The most common form of error in task A was classed under the 'other errors' (i.e. not preservative or set-loss errors). This indicates that the participants were engaging in a process of trial and error in order to establish what the correct sorting criterion was. The second most common form of error was the preservative error, which also occurred quite frequently. This indicates that the participants failed to adapt when the sorting criteria changed and stuck to the previous rule, despite being informed that they were incorrect. Very few set loss errors were committed, indicating that understood when they chose the correct criterion and were able to maintain that rule until the criterion changed. Correct reaction times of participants also decreased through the trials in task A, indicating that the participants were developing a better understanding of the task.

These behavioural results are largely in line with those of Lie et al. (2006). They also found that the 'cognitive costs' were highest for A, with a decrease in task complexity associated with a fall in cognitive costs. The level of errors in this study also match with those of Lie et al. (2006), who reported that task A was associated with the highest level of errors, with the number of errors falling in tandem with task complexity. Also in line with the findings of Lie et al. (2006), the largest number of errors for task A was 'other' errors rather than preservative or set-loss errors. This demonstrates that the participants effectively engaged in the tasks.

Inspection of the fNIR data further supports the design of the task and the difficulty hierarchy. Task A produced significantly higher blood oxygenation values that the HLB(A). This indicates that the full WCST resulted in increased neural activity in the PFC. This is in line with neuroimaging and lesion studies that have demonstrated the importance of the PFC for the WCST and gives us confidence that our fNIR device is successfully measuring blood oxygenation in the DLPFC. The fact that task B did not produce significantly higher levels of blood oxygenation than HLB(B) was to be expected as the participants found task

B to be relatively easy. The fNIR results also support previous neuroimaging and lesion studies, which have reported that areas of the PFC, including the DLPFC are utilised during the WCST (Haines, 1994; Milner, 1963; Rezai et al., 1993). One difference between the behavioural and fNIR results is that the significant differences found between tasks A and B in the behavioural data were not replicated in the fNIR data. A possible explanation is that the fNIR device is dependent upon relatively large differences in task difficulty in order to record significant differences in the change in blood oxygenation. The fact that significant differences in blood oxygenation were found when there were large differences in task difficulty (A compared to the HLBA) but not when task difficulty was closer (A compared to B and B compared to the HLB) would support this explanation. Overall, the fNIR data and behavioural data suggest that increasing task complexity, due to cognitive switching, had a significant effect on task performance and DLPFC blood oxygenation.

One consistent effect found in the fNIR results is that the superiority of voxels was related to differences in the bloody oxygenation, with inferior voxels displaying greater levels of oxygenation. An effect of laterality was also found, with greater levels of blood oxygenation in voxels on the left. Previous research on the WCST has reported that the WCST activates the left PFC (Haines, 1994), whereas other research has reported that it activates that right PFC (e.g. Volz, 1997). As previous research has been very inconsistent with regards to the lateralisation of WCST neural activity, it is hard to know how to interpret these results.

5.6.2 CBT

The behavioural results are largely in line with predictions, showing that the participants found the BST most difficult, followed by the CBT and then the baseline task. This gives support for the design of the study that was designed to have a difficulty hierarchy of BST>CBT>Baseline. The proportion of correct responses for the BST was significantly worse than for the CBT, which were in turn significantly worse than for the baseline. Sequence length also had an affect on the proportion of correct responses, with a significant decrease in the proportion of correct responses for the BST and CBT as the sequence length increased from

3-6 blocks. An interaction revealed that at sequences lengths of 5 and 6, performance on the BST and CBT was significantly worse than for the baseline task as well as performance for the previous sequence lengths. This demonstrates that the differences between the tasks emerged when the sequence lengths increased. This is presumably due to the CBT and BST becoming more difficult with increasing sequence length, as the tasks require more complex spatial working memory and inhibitory functions. At a sequence length of 5 the BST and CBT were significantly different, with fewer correct responses for the BST.

Reaction times for correct responses largely mirror the results from the proportion of correct responses. Correct reaction times were slower for the CBT and BST than the baseline, indicating that the participants found the CBT and BST more difficult than the baseline. Correct reaction times were also affected by sequence length, with increasing reaction times due to increasing sequence length. This indicates that increased sequence length affected task difficulty and performance. An interaction between task and sequence length revealed that the CBT and the BST were significantly different from the baseline at sequence lengths 5 and 6, with significantly slower reaction times at these sequence lengths. This indicates that increasing sequence length increased task difficulty for the CBT and BST but not the baseline. Increasing sequence length also revealed differences between CBT and BST, with slower reaction times for the CBT at sequence lengths of 3 and 4. Overall, the reaction times again indicate that the BST was the most difficult task, followed by the CBT, with the baseline task being the easiest.

The behavioural results from this study are in line with those of Toepper et al. (2010) on which this study is based. They found that as task difficulty (BST>CBT>Baseline) and sequence length increased, the number of errors and reaction times increased. Given that the average spatial span of individuals is 5 (Kessels et al., 2000), it is not surprising that task performance began to fall significantly at this point. When dealing with sequence lengths of 5 or more, maintaining this spatial information in working memory becomes increasingly difficult, especially for the BST where the participants also have to inhibit supress the irrelevant spatial information from the distractor blocks. Combined, the behavioural data suggests that increased spatial working memory and inhibitory demands from the tasks impacted on performance.

The behavioural results from the CBT, BST and baseline tasks are in line with those of Topper et al. (2010), who found that reaction times increased from the baseline, to the CBT and then the BST. Just like in this study, Topper et al. (2010) also found that reaction times increased as the sequence length increased. The finding that increasing sequence length affected the BST and CBT but not the baseline and that the difference between the CBT and BST varied with sequence length also replicated the findings of Toepper et al. (2010). The accuracy findings of this study mirror those of Toepper et al. (2010). As in the Toepper et al. (2010) study, errors were significantly higher in the CBT and BST than in the baseline, while increasing sequence length was associated with an increase in errors for the CBT and BST but not the baseline task. The behavioural results from the CBT condition are in agreement with the standard literature on the CBT, with accuracy falling as the sequence length is increased (Kessels, Zandvoort, Postma et al., 2000).

Despite the significant differences in the behavioural measures, no such differences were found in the fNIR data. The behavioural results suggest that the study was designed and executed in a proper manner. The design of the CBT study as based on that of Toepper et al. (2010), and while the behavioural results a largely in line with theirs, the neuroimaging results are not. Previous neuroimaging the lesion studies have strongly indicated that the CBT, as well as the BST, activate areas of the PFC, in particular the DLPFC. One possible explanation for this discrepancy is relates to the nature of VSMW. While the PFC has been found to be involved in VSWM, it is not specific to this area. Research has shown that areas such as the parietal cortex and the hippocampus are also involved in VSWM (Toepper et al., 2010). Meanwhile neural activity related to the WCST and COWAT appears to be more localised (Alvarez and Emory, 2006).

This links to a further issue, which is the different neuroimaging methods used. While this study used fNIR, many of the previous neuroimaging studies on the CBT have used fMRI or PET. One of the advantages techniques such as fMRI have over fNIR (the system used in this study specifically) is the ability to record neural activity over larger areas and greater depths of the brain. Not only is the fNIR device incapable of recording activity in areas such as the parietal cortex, it may also struggle to record activity in areas such as the VLPFC due to the positioning of the sensor pad on the forehead and the limited depth sensitivity of the device (this is why the device is best used for recording DLPFC activity). This may be a problem given that research has shown that the VLPFC is activated during the CBT (Owens et al., 1993, 1996; Toepper et al., 2010).

Another possible explanation relates to the recording periods used in the CBT tasks. The window of activation recorded for each task and sequence length is smaller than those used in the WCST and the COWAT, with a 21 second window of activation for a sequence length of 6 (the maximum sequence length used). This includes both the encoding and response phases. However, Toepper et al. (2010) suggested that frontal lobe activity during the CBT and BST is associated with the encoding phase, meaning that the response phase may be a confounding factor. Reducing the window to only include the encoding phase introduces increased noise and variance into the signal, meaning that the fNIR device may not have had the required signal-to-noise ratio to record the haemodynamic activity. When looking at the fNIR results for the CBT, it does appear that the results are heading in the predicted direction, with an increase in blood oxygenation from the baseline to the CBT and then the BST. However the effect was not pronounced enough to be significant. This may be due to the physiological or methodological issues mentioned above.

5.6.3 COWAT

When examining the number of correct words generated in the generation condition, it is clear that there was a wide range in the number of words produced by the participants, with one participant managing only nine correct words while another managed to produce fifty-seven correct words. A possible explanation for this result is that the participant who generated the fifty-seven correct words was studying for an Undergraduate degree in English Literature. As a result that participant may have had a more extensive lexicon, allowing them to generate such a large number of words on sixty seconds. This is in line with research that has suggested that a person's reading ability affects their verbal fluency performance (Borkowski et al. (1967).

Finding results in the literature with which to compare the behavioural results from the COWAT is difficult due to the COWAT predominantly being used to test clinical populations or specific demographics, as well as the fact that multiple sets of letters have been used in previous studies. Tombaugh, Kozak and Rees (1999) reported that mean number of correct words produced in a COWAT was 38 to 44 for 16-59 year olds depending on their level of education. Ruff, Light and Parker, (1996) found that the average number of words produced by people of all education levels was 40.1. Meanwhile Ross (2003) found the average number of words produced by undergraduates was 38.16. Ross (2003) also found that the mean number of clusters was 7.7, the mean cluster size was 0.42 and the total number of switches was 24.8. Ross, Calhoun, Cox et al. (2007) reported that average number of words produced by undergraduates in a COWAT was 37.5. They also reported that mean number of clusters was 7.7, the mean number of switches was 22.8 and the mean cluster size was 0.41. The results from this study are quite a lot lower in terms of the average number of words produced and the number of switches, while the number of clusters is similar and the average cluster size is larger. However, a study by Troyer (2000) reported that the average number of words produced in COWAT by 18 to 91 year olds was 28.6, with age having little effect on the number of words produced. This figure is much closer to the average of 26 words produced in this study. However, the mean number of switches was again higher and the mean cluster size lower than in this study.

One explanation for these differences relates to the design of the COWAT used in this study. Standard versions of the COWAT use three letters for word generation whereas this study used two, as well as two letters for reading. Given the ABBA/BAAB design of this study, as well as the addition of the letter 'R' that is not used in standard versions of the COWAT, the participants will only have produced words using two of the letters used in standard versions of the COWAT (Spreen and Strauss, 1998). A further explanation may relate to sample size. 27 participants were tested in this study whereas many of these previous COWAT studies have used hundreds of participants (e.g. Tombaugh, Kozak and Rees, 1999). It may be that the comparatively small sample size in this study contributed to the relatively low number of words produced. It should be noted that while the behavioural results are lower in comparison to the previous results detailed above, they are still within the region of these previous COWAT results and the relationship between the different recorded measures is similar to those reported.

In line with predictions, the generation task was associated with significantly higher levels of blood oxygenation than the reading task, indicating that the verbal fluency task resulted in increased neural activity in the DLPFC compared to reading aloud. This supports previous neuroimaging and lesion studies that have found increased activity in the DLPFC during the COWAT (Frith, 1995; Warkentin and Passant, 1997). No effect of laterality was found, which is perhaps surprising given that a number of neuroimaging and lesion studies have reported that the COWAT performance is associated with the left DLPFC (Frith, 1995; Warkentin and Passant, 1997; Baldo et al., 2001). There was however a significant effect of superiority, with greater levels of oxygenation in inferior voxels. There was also a significant correlation between the four voxel areas and the mean cluster of words generated, indicating that the more words the participants generated the higher the levels of blood oxygenation. The changes in blood oxygenation throughout the duration of the generation and reading tasks (as seen in Figure) reveal that blood oxygenation increases throughout the generation task and falls during the reading task. As the participants progressed through the generation task it would have become increasingly difficult for them to generate correct words without producing homophones and repetitions. This would presumably have resulted in increased task cognitive demand and task difficulty, which may explain the increasing levels of blood oxygenation. Meanwhile in the generation task, blood oxygenation may have fallen as the task progressed as the task difficulty did not increase. As a result the participants may have become accustomed to the task early on and as a result the blood oxygenation levels in their DLPFC fell.

Overall, the fNIR and behavioural data indicate that the process of accessing the mental lexicon, searching for appropriate words and switching between clusters of words produces greater levels of cognitive demand and blood oxygenation than simply reading words aloud.

5.6.4 General

The results of the WCST and COWAT suggest that the fNIR device is sensitive to the haemodynamic activity associated with these tasks as the device recorded increased blood oxygenation during these tasks in comparison to control tasks that were less cognitively demanding. This is in line with previous neuroimaging research and lesion studies, which suggest that this haemodynamic activity is primarily occurring in the DLPFC, which in turn suggests that the fNIR device is sensitive to haemodynamic activity in the DLPFC. This is also in line with the predictions for the study. The failure to find any significant haemodynamic effects in the CBT tempers this though. It is possible that this failure was related to the relatively short phases of the CBT tasks, or that the spatial working memory utilised during the tasks was primarily activating non-frontal areas (i.e. parietal cortex and hippocampus), or frontal areas that the fNIR device did not posses the ability to measure (i.e. the VLPFC). Compared to the CBT, the WCST and the COWAT display much more localised brain activity (Alvarez and Emory, 2006). This opens up the potential that the CBT and BST were activating areas of the brain that the fNIR device could not record activity from.

A common requirement in both the WCST and the COWAT is the need to inhibit inappropriate responses. In the WCST, respondents need to inhibit responding on the basis of a previous sorting criterion. In the COWAT, respondents need to inhibit the generation of words that have already been produced as well as any homophones. Although not significant, the CBT data indicated that blood oxygenation was greatest during the BST. Compared to the CBT, the major difference in the BST is the need to inhibit and supress irrelevant spatial information. It is possible that the increased blood oxygenation in the tasks was at least in part due to task inhibition. As detailed in Chapter 4, the DLPFC is heavily involved in the process of inhibition, which supports the possibility that the increases in haemodynamic activity during these tasks were due to task inhibition.

Another factor that may explain increased blood oxygenation during the tasks is task workload. The WCST and COWAT tasks that produced the highest levels of blood oxygenation were also the tasks that contained the highest levels of task workload. Task A in the WCST and the word generation condition in the COWAT ultimately required increased workload to complete the tasks than the other tasks in their respective experiments due to the increased difficulty and demands of the tasks. And although not quite significant, the task that produced the highest levels of haemodynamic activity in the CBT task was the BST, which involved the greatest task workload due to its increased difficulty and task demand.

Research has revealed that task workload affects activity in the DLPFC and that fNIR is able to record workload levels through the measurement haemodynamic activity in the DLPFC. Herff et al. (2014) used fNIR and an n-back task to show that fNIR was able to accurately quantify mental workload in the PFC, with increased task workload leading to increased haemodynamic activity in the PFC. Ayaz et al. (2012) demonstrated that fNIR was able to accurately record levels of mental workload in ecologically valid environments. They found that fNIR was able to record levels of mental workload in the DLPFC during air traffic control operations by experienced air traffic controllers, with increased task workload associated with increased DLPFC activity. Further evidence of fNIR recording levels of workload in the DLPFC comes from Izzetoglu et al. (2003) who analysed haemodynamic changes in the DLPFC during the 'Warship Commander Task'. They found that changes in task workload had significant effects on blood oxygenation in the DLPFC. It is therefore possible that frontal lobe activity in the experimental tasks was at least in part related to task workload.

In summary, fNIR appears to be able to reliably record haemodynamic changes in the DLPFC, with inhibition and workload being the potential sources of the haemodynamic activity recorded by the fNIR device.

Chapter 6 – fNIR and Simulated Driving

6.1 Abstract

Functional near-infrared spectroscopy (fNIR) provides a means with which to examine the cognitive functions associated with various aspects of driving. So far a limited amount of driving research has investigated the neuroscience of driving. This study utilised fNIR in order to investigate the potential role of the dorsolateral prefrontal cortex (DLPFC) in driving. Participants took part in four simulated driving overtaking tasks that varied in the levels of workload and inhibition required. Subjective workload was recorded along with fNIR recordings of DLPFC activity during the driving tasks. The results revealed differences between DLPFC activity and subjective workload between the different driving tasks. Two factors could potentially explain the results of the experiment, task workload and inhibition. The relative merits of these explanations and future research are discussed.

6.2 Introduction

Study 3 revealed that fNIR was able to reliably record DLPFC activity. The study also indicated that fNIR was recording DLPFC activity associated with workload and inhibition. The next step is to record frontal lobe activity with fNIR in a controlled driving environment in order to establish what role the DLPFC play in driving. Particular focus will be given to task workload and inhibition as the frontal lobes are involved with these functions and fNIR appeared to be recording these functions in the previous study.

6.2.1 Workload and Driving

As detailed in Chapter 4, previous driving research that has used fNIR has reported that task workload, in particular task difficulty, affects DLPFC activity during driving.

6.2.2 Inhibition and Driving

A major role of the frontal lobes is to exert executive control (Badre and D'Esposito, 2009). A method of exerting executive control over the brain is through cognitive inhibition. Research has shown that the executive control and inhibition exerted by the frontal lobes is key to decision-making and impulse control (Bechara and Van der Linden, 2005; Krawczyk, 2002). Inhibition and impulse control may play an important role in driving. For example, a lack of the inhibition and impulse control provided by the frontal lobes may lead to drivers speeding, pulling out at roundabouts or overtaking other cars when it is not appropriate or safe to do so. Krawczyk (2002) reported that an important function of the DLPFC is to combine multiple sources of information in order to guide behaviour and actions. As safe driving is a process of using multiple sources of information to guide driving behaviour, it is important to explore the role that the frontal lobes play in driving.

As mentioned in Chapter 4, numerous studies have shown that one of the major functions of the frontal lobes is to provide inhibition. The role of inhibition is to prevent behaviours that may be detrimental and, in terms of driving, such behaviours may have very serious consequences. Research has also shown that inhibition plays an important role in driving. There has been particular focus on how a lack of inhibition is involved in unsafe driving.

O'Brien and Gormley (2013) investigated the role that impulsivity and inhibition play in the dangerous driving of young people. They compared the performance of young drivers who had committed speeding offences with non-offenders in Stop-signal and Go/no-go tasks (both of which are inhibitory measures). While no differences between the two groups were found in the Stop-signal task, significant differences were found with the Go/no-go task. Offenders were found to exhibit lower inhibitory skills on the Go/no-go task as they made more errors on the Go tasks than non-offenders. However, offenders also responded more quickly on the Go tasks. The authors interpreted this as being the result of a speed-accuracy trade-off. This research suggests that a lack of inhibitory abilities may be a factor dangerous driving and road offences.

Research has also focused on how alcohol increases dangerous driving behaviours by diminishing the ability to inhibit risky behaviours. Fillmore, Blackburn and Harrison (2008) studied the effect that alcohol had on risky driving behaviour. The researchers wanted to investigate how the inhibitory impairing effects of alcohol can be exacerbated by situations of conflict in which the inhibition and expression of dangerous behaviours are equally motivating. They had subjects complete a go/no-go task to measure inhibitory control and a simulated driving task to evaluate driving behaviour. Driving performance was gauged using measures such as lane deviations, vehicle impacts and stopping at traffic lights. The participants were provided with monetary incentives for these tasks for slow and careful behaviour (e.g. slow driving and inhibiting impulsive behaviour), as well as for fast and abrupt behaviour (e.g. fast driving and disinhibition) in order to produce conflicting motivations. They reported that alcohol consumption interacted with conflict to diminish inhibitory control and to increased dangerous driving. The participants who displayed the worst driving were the participants whose inhibitory control had been most affected by the alcohol, demonstrating that inhibition is vital for safe driving.

Cheng and Lee (2012) investigated the relationship between response inhibition, risk-taking and the driving behaviour of motorcyclists who had different levels of impulsivity. Response inhibition was measured using a Stroop test, impulsivity was measured using the Barratt Impulsiveness Scale, risk taking was assessed using the Balloon Analogue Risk Task and risky driving behaviour was measured with a Rider Driving Violation scale (which consisted of a self-report scale on driving violations committed). The results of the study revealed a significant relationship between response inhibition, impulsivity and risk taking. Motorcyclists with higher levels of impulsivity displayed lower levels of response inhibition and greater risk taking. Motorcyclists with high or medium levels of impulsivity were five times as likely to be involved in accidents as those with low impulsivity. This study demonstrates that drivers with higher levels of response inhibition are safer drivers and are less likely to be involved in crashes.

Cheng, Ng and Lee (2012) investigated whether motorcyclists with a history of driving offences would make riskier driving decisions and engage in more risk taking as a result due to poor response inhibition and impulsivity. Motorcyclists who had been responsible for an accident and who had at least 6 points on their licence were compared to motorcyclists who had not committed any offences. Response inhibition, impulsivity and risk taking were measure in the same way as the Cheng and Lee (2012) study while risky decision-making was measured using a simulated motorcycle Go/no-Go task. In this task the participants had to decide whether or not to abruptly switch from the right lane to the left lane of a roundabout (which was occupied by a bus) in order to make a left turn. The authors reported that the motorcyclists with a history of offences differed in terms of response inhibition, risk taking and risky driving. The offenders displayed worse response inhibition and a greater willingness to take risks. They did not differ in terms of impulsivity though, with the authors attributing this to problems with the self-report method used. Again, this study provides evidence that levels of response inhibition are associated with how safe and accident-prone drivers are.

As mentioned in Chapter 1, one of the major findings in driving research is that young drivers are overrepresented in crash and fatality statistics. Research has also shown that adolescents and young adults lack inhibition, are impulsive and they engage in more risk taking and dangerous behaviours (Weinberg, Elvevag and Giedd, 2005; Steinberg 2008). It is possible that the increased crash and fatality rates of young drivers is in part related to their lack of inhibitory abilities, which may in turn indicate the role that inhibition plays in driving.

Together, these studies reveal the important role that inhibitory control plays in driving. High levels of response inhibition are associated with safer driving, reduced likelihood of being involved in an accident and the lower likelihood of committing traffic offences. Conversely, poor response inhibition is associated with dangerous driving, a higher likelihood of being involved in an accident and a higher likelihood of committing traffic offences. So far there have not been any studies that have used NIRs to investigate inhibition in driving.

6.3 Study 4

Study 4 will consist of a simulated overtaking and following experiment that uses fNIR to monitor DLPFC activity. Frontal lobe haemodynamic activity will be monitored with fNIR throughout 4 driving tasks that vary in levels of inhibition and workload and then compared between the conditions. Overtaking and following tasks was chosen because they are driving tasks that allow the manipulation of task workload and inhibition. As mentioned previously, both workload and inhibition have been reported to be involved in driving and frontal lobe function, while Study 3 indicated that fNIR was recording these functions. Clarke, Ward and Jones (1998) reported that the most important factor in a safe overtaking manoeuvre is inhibiting the desire to overtake when it is not safe. This suggests that an overtaking task may be able to elicit inhibitory responses while driving. Meanwhile previous driving studies (detailed in chapter 4) using fNIR have found that changes in DLPFC activity are primarily related to the workload associated with driving. A following task was chosen to provide similar levels of workload (the participants still have to drive in the simulator) as the overtaking task, but not the same levels of inhibition. In the following task, the participant will still be required to drive the same simulated driving route; the difference is that they will not be instructed to overtake any vehicles.

The experiment will be further divided into 'active' and 'passive' conditions. In the active condition, the participants will drive in the simulator. In the passive conditions the participants will not drive, but instead watch replays of the following and overtaking tasks other participants have completed in the active condition. The aim of the passive tasks is to test whether any differences in frontal lobe activity are the result of visual stimulation by providing the same visual stimuli without the workload and inhibition of the active tasks. This method also provides a further workload comparison as the active tasks involve the workload of driving in the simulator while the passive tasks do not.

Subjective workload scores will be recorded with the NASA-TLX workload scale in order to gauge the general levels of workload the participants experience in the different driving tasks. This may help provide insight into the fNIR data regarding the role of workload in DLPFC activity. The NASA-TLX scale was chosen because it is a quick and easy to use measure of subjective workload, and Tsunashima and Yanagisawa (2009) used it with success in a previous fNIR driving study (finding a strong correlation between HbO_2 levels and NASA-TLX workload scores).

The aim of study 4 is to establish if fNIR is capable of reliably recording DLPFC activity during simulated driving, as well as seeing if fNIR is a suitable device for using in simulated driving environments. A further aim of this study is to establish what role the DLPFC may play in driving. In particular, are any inhibitory functions associated with the DLPFC involved in driving or is their primary function related to task workload? The predictions for Study 4 are that the active tasks will result in increased DLPFC activity and subjective workload levels compared to the passive tasks. It is also predicted that the overtaking tasks will result in increased DLPFC activity and subjective workload levels compared to the following tasks. The final prediction of this study is that the highest levels of DLPFC activity and subjective workload will be found in the active overtaking task.

As in study 3, oxygenation values will be used to measure DLPFC activity. However, this study will also report separate Hb and HbO_2 values. The reason for this decision is that a number of previous driving studies that have utilised fNIR have reported a wide range of measures including blood oxygenation, Hb and HbO_2 whereas previous lab studies (most similar to those used in study 3) focused on blood oxygenation.

6.4 Methodology

6.4.1 Participants

A power analysis revealed that for a large effect size (F = 0.4), a total of 15 participants would be needed for this study.

In total, 19 participants were recruited for the study (13 female, 6 male). All of the participants were students at the University of Nottingham and had normal or corrected-to-normal vision. Data from 3 participants was removed from the analysis due to missing data. This left a total of 16 participants (11 female and 5 male).

The mean age of the participants was 23.75 years with a range of 19 to 27 years. Mean driving experience was 4.75 years of driving with a rage of 8 months to 7 years. Mean number of miles driven per year was 5605.625 with a range of 20 to 10,000 miles.

6.4.2 Apparatus and Stimuli

6.4.2.1 Driving Simulator

The experiment took place in the Nottingham Integrated Transport and Environment Simulation Facility (NITES), with the NITES 2 driving simulator selected for use in this experiment. The NITES 2 driving simulator consists of a stripped-down driving rig and a 180 degrees of vision projection screen (Figure 1). The rig consists of a seat, a steering wheel and a gear lever integrated together (Figure 2). Three projectors are used to present a simulated driving environment on the projection screen.

6.4.2.2 Simulated Driving Scenarios

Custom made driving simulation software (XPI Simulation, UK) was used to create a simulated driving environment and specific tasks (scenarios) within that environment. The driving environment itself consisted of a 3-mile single-carriageway road (as seen in Figure 1) with hedges, trees, fields and some buildings appearing on each side of the road. The single-carriageway road had a speed limit of 60 mph and therefore took 3 minutes to travel along if driving at a consistent speed of 60mph. The traffic density on the road was such that approximately 40 vehicles would drive past in the opposite lane during the driving scenario if the participants drove at the speed limit of 60mph. This traffic density was mirrored in the lane that the participants drove in. The traffic density resulted in the opposite lane being clear for approximately 50% of the drive. The vehicles travelling in the same lane as the participant's simulated vehicles were programmed so that they drove at a speed of 50mph. This was done in order to help facilitate overtakes that did not require the participants to break the speed limit. The four tasks used in this study were:

- An active overtaking task in which the participant was instructed to overtake as many cars as they thought safe during the 3-mile journey.
- An active following task in which the participant was instructed not to overtake any other vehicles but instead to follow the vehicle ahead.

- A passive overtaking task in which the participant watched a video of a pre-recorded overtaking task in the simulator.
- A passive following task in which the participant watched a video of a pre-recorded following task in the simulator.

Each one of these driving tasks involved simulated driving scenarios that took place in dry, daytime road conditions. The passive overtaking and following tasks took approximately 3 minutes each to complete whereas the length of the active overtaking and following tasks was dependent on the speed at which the participants drove. The participants were instructed to try and drive at the speed limit to ensure that they all completed the tasks in roughly the same amount of time. The active overtaking task has been designed to require high levels of inhibition (from the overtaking manoeuvres) and workload (resulting from the general driving in the task). The active following task will require similar workload demands, as the participant still has to drive the car in the simulation, but will not elicit the same levels of inhibition. The purpose of the replay tasks is to remove the workload from the tasks, as the participant does not have to drive, and see what effect this has on DLPFC activity compared to the active tasks. The replay tasks also present the same visual stimulus as in the active tasks in order to rule out the possibility that increased frontal lobe activation could simply be a result of visual demand.

From the driver's perspective, they see the road ahead but they do not see any part of the simulated car that they are operating. The simulated car itself occupies the same area of space on the road as a real car would on a single-carriageway road. The driver is essentially positioned on the driver's side with an invisible passenger side located next to them.

Figure 6.1. View of the NITES 2 driving simulator with the driving rig (centre) and fNIR device (right).

Figure 6.2. NITES 2 driving rig and a participant with the fNIR sensor on their forehead.

6.4.2.3 fNIR Device

The same fNIR device that was used in Study 3 was also used for this study. For further information on the fNIR device, please refer back to chapters 4 and 5.

6.4.2.4 NASA TLX

The NASA TLX (Hart and Staveland, 1988) workload scale was used to assess the workload associated with each of the driving tasks. 6 scales are used to measure workload. The scales are Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. Each scale is scored out of 100, with higher scores being an indicator of higher workload. For the NASA-TLX scale see Appendix 6.1.

6.4.2.5 Simulator Sickness Questionnaire

A standardised simulator sickness questionnaire was used to assess simulator sickness levels in the participants. It can be found in Appendix 3.1.

6.4.3 Design

A repeated-measures design was utilised in this study, with all of the participants completing each of the four driving tasks. The dependent variables were the workload scores for the 6 NASA-TLX scales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration) and the changes in Hb, HbO_2 and blood oxygenation. The independent variables were task (following or overtaking) and involvement (active or passive).

6.4.4 Procedure

The participants were first shown the driving rig and had the controls of the driving rig explained to them. They were then given a 1 minute practice drive (along an empty single-carriageway road) to allow them to familiarise themselves with the driving simulator and its controls.

Once the practice drive had been completed and the participant was comfortable with the simulator controls, the fNIR sensor was secured on their forehead (making sure that there was no hair in the way) using Velcro straps and a cloth (to keep out external light). An fNIR reading was then taken to make sure that a clear signal was being received. If this was not the case, the sensor pad was re-placed on the forehead as a bad signal often results from the sensor pad not being firmly secured to the forehead. Once a good signal was achieved, a baseline reading (taking 10 seconds) was taken while the participant stared at the blank projection screen.

When the participant was comfortable driving in the simulator, the fNIR sensor pad was placed on their head and secured in place with Velcro straps and a bandage (which also helped to keep out external light). Brief fNIR recordings were then taken in order to establish the quality of the fNIR signal being received. If the signal was saturated or low, the sensor position was adjusted on the participant's forehead to improve contact and the drive current and gains of the fNIR device were also altered if necessary. When the fNIR signal was within the acceptable boundaries (free of saturated channels and low level channels) the settings were left in place for that participant.

Before each driving task the participant was given an explanation as to what the task was and instructed on what they needed to do. For the active overtaking task the participants were instructed to overtake as many vehicles as they could while obeying traffic laws and driving in a safe manner. For the active following task the participants were instructed to follow the vehicle in front, while keeping a safe and constant distance, for the duration of the simulation. The participants were informed of the speed limit of the road (60mph) and instructed to try and drive at the speed limit. For the passive following and overtaking tasks the participants were instructed to simply watch the pre-recorded driving simulations that were projected in the simulator. An fNIR baseline was then taken and the participant was told to relax and look straight ahead at the projection screen (which was just blue at this point with no images from the driving scenario being projected). The baseline process took approximately 10 seconds. Once the fNIR baseline was taken, the driving scenario was activated and the fNIR was set to record frontal lobe activity.

After each hazard scenario had been completed a short break was taken during which time the participants completed the simulator sickness questionnaire (Appendix 3.1). If at any point scores of ‘Moderate’ sickness were reported then the experiment was paused and the participant was provided with some water. The participants were given several minutes after which they were asked if they wanted to continue. If they said that they felt unable to continue then they were removed from testing. If the participants stated that they now felt O.K. then the experiment continued. If at any point a participant reported a score of ‘Severe’ on any of the measures then they were provided with some water and they were removed from testing. A fan blowing cool air was left on at all times to help reduce incidences of simulation sickness. No participants were removed from testing due to simulator sickness in this study.

The NASA TLX workload questionnaire was also completed after each task. This resulted in a gap of 2 minutes between one task finishing and the next task starting. The same procedure was completed for each of the 4 driving tasks. Each of the passive tasks lasted approximately 3 minutes, whereas the active driving tasks lasted approximately 3 minutes if they drove at the speed limit of 60mph. This meant that if each participant drove at an average of 60mph in the active conditions then the four driving tasks would have lasted for approximately 12 minutes in total. During the active tasks, average speeds ranging from 50mph to 60mph were observed. The order of the tasks was arranged so that no participant completed two overtaking or following tasks in succession.

If the participants crashed at any point during the active following or overtaking tasks then the scenario was restarted unless the crash occurred very close to the end of the simulation.

6.5 Results

6.5.1 NASA TLX Data

Data from the NASA TLX workload scales were first analysed using a 2x2x6 repeated-measures ANOVA. The factors were: Involvement (with the levels Active and Passive), Task (with the levels Following and Overtaking) and Scale (with the levels: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration). Pre-processing of the data involved taking the square root values of the scale scores in order to avoid breaking the homogeneity of variance assumption. Mauchly's Test of Sphericity was breached for the interaction between Involvement and Scale as well as the Task and Scale interaction. Greenhouse-Geisser corrections were therefore used to calculate the degrees of freedom for these values.

Six further 2x2 repeated-measures ANOVAs were used to analyse data from each of the 6 scales separately. The factors were: Involvement (with the levels Active and Passive), Task (with the levels Following and Overtaking).

The results of the ANOVA revealed significant main effect of Involvement $F(1,15) = 340.651$, $MSE = 144.451$ $p<.001$, ${}_{p}\eta^2 = .958$ with NASA TLX scores generally higher in the active (driving) conditions compared to the passive (replay) conditions. There was also a significant main effect of Task $F(1,15) = 90.365$, $MSE = 57.165$, $p<.001$, ${}_{p}\eta^2 = .858$ with workload scores being higher in the overtaking tasks compared to the following tasks. A significant interaction between Involvement and Task was found $F(1,15) = 95.911$, $MSE = 95.911$, $p<.001$, ${}_{p}\eta^2 = .865$, which is shown in Figure 6.3.

A significant main effect of Scale $F(5,75) = 9.372$, MSE = 3.145, $p<.001$, ${}_{p}\eta^2 = .385$ demonstrated that scores on some of the workload measures were generally higher than others. There were also significant interactions between Involvement and Scale $F(2.619, 39.283) = 3.495$, $MSE = 1.994$, $p<.05$, ${}_{p}\eta^2 = .189$, as well as between Task and Scale $F(3.496, 52.447) = 5.659$, $MSE = 1.573$, $p<.05$, ${}_{p}\eta^2 = .274$. However, the three-way interaction between Involvement, Task and Scale was not significant [$F(5, 75) = 0.980$, *MSE*.171, $p = .415$, ${}_{p}\eta^2 = .061$], implying that the interaction in Figure 6.3 was broadly similar for all scales. Table 6.1 illustrates the mean scores and effect sizes for the 6 workload scales when separated by Involvement and Task, from which it can be seen that while both main effects are significant for all six scales, the effect size is largest on the Mental Demand and Effort scales.

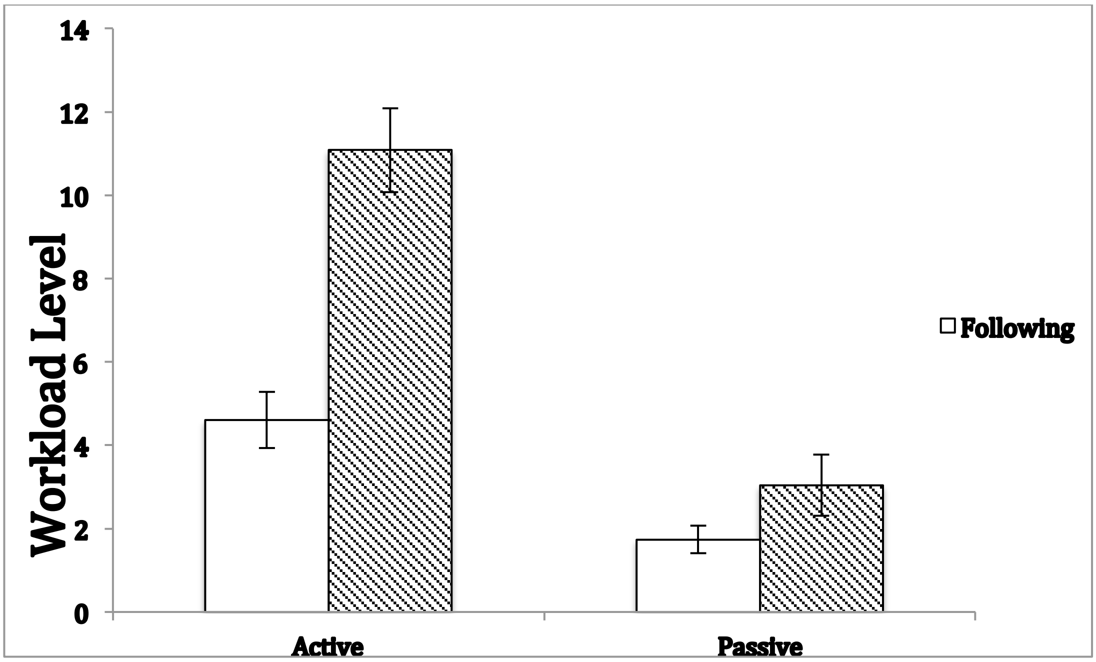


Figure 6.3. Significant interaction between Involvement and Task on workload scores. Error bars represent +/- 1 S.E. *Raw NASA TLX values were used to create this figure.*

*Table 6.1. Descriptive statistics of the 6 scales of the NASA TLX workload measure. Partial eta squared values are calculated from the main effects from six individual 2x2 ANOVAs. When p<.05, p = *; when p<.01, p = **; when p<.001, p = ****

		Involvement				Task			
TLX Scale		Active	Passive	pEta^2	p	Follow	Over.	pEta^2	p
Mental Demand									
	mean	3.025	1.644	.927	***	1.828	2.841	.806	***
	sd	0.866	0.636			0.697	1.059		
Physical Demand									
	mean	2.328	1.198	.802	***	1.509	2.017	.716	***
	sd	0.794	0.258			0.480	0.996		
Temporal Demand									
	mean	2.679	1.527	.818	***	1.535	2.672	.733	***
	sd	0.937	0.748			0.584	1.056		
Performance									
	mean	2.329	1.278	.795	***	1.558	2.049	.606	***
	sd	0.831	0.297			0.548	0.961		
Effort									
	mean	2.946	1.278	.940	***	1.759	2.465	.825	***
	sd	0.769	0.532			0.763	1.213		
Frustration									

mean	2.662	1.685	.624	***	1.786	2.561	.588	***
sd	0.847	0.652			0.694	0.919		

6.5.2 fNIR Data

Pre-processing of the fNIR data involved averaging the Hb and HbO_2 values for each of the 16 channels over time (the entire duration of the task). The fNIR data was averaged over time in order to get an overall, global view of the changes in haemodynamic activity throughout the driving tasks. During the driving tasks there were multiple points of interest and differences in behaviours and processes required. The major difference between the active and following tasks related to the difference in motor commands. In the passive tasks the participants did not have to drive and operate the simulated car controls. The major difference between the active following and active overtaking tasks related to decision-making. In the active overtaking task the participants simply had to follow the car in front. In the active overtaking task the participants had to engage in a decision-making process every time they attempted to overtake. There was also a difference in motor control between the active tasks, with overtakes or attempted overtakes requiring precise and accurate control inputs compared to the simple and constant control inputs needed for the following task. These differences would have been present throughout the duration of the driving tasks. It was decided that encapsulating all of these factors into the analysis was important. Providing individual analyses for all of these events would not have been practical.

The decision to average over time was also made as the experimental setup made it very difficult and time-consuming to provide accurate synchronisation between events in the simulator and the fNIR data. There were also difficulties in finding comparable events that could be studied across all or most of the participants. For example, participants differed significantly in the number of overtakes they made or attempted, as well as the time they took to do these things. To balance this out, a further analysis that focused on individual points of interest was also made. In order to remove any erroneous data from 'bad' channels, data that fell beyond 2 standard deviations (positive or negative) was removed and replaced with the participant's global condition mean. For Hb, 4.17% of values were replaced. For HbO_2, 3.74% of values were replaced. In order to calculate blood oxygenation values, Hb values were subtracted from HbO_2 values for each of the 16 fNIR channels. These values were also averaged for each of the 16 channels over the duration of the task. All of the blood values are changes relative to baseline and are measured in micro moles (μM).

Oxygenation, Hb and HbO_2 values were each analysed separately using a 2x2x16 repeated measures ANOVA. In all of the ANOVAs the three factors were: Involvement (with the levels Active and Passive), Task (with the levels Following and Overtaking) and Site (with 16 levels corresponding to the 16 fNIR channels).

6.5.2.1 Oxygenation Values

For the 2x2x16 ANOVA of the oxygenation values, Mauchly's Test of Sphericity was breached for Site, the Involvement and Site interaction, the Task and Site interaction and the three-way interaction between Involvement, Task and Site. Greenhouse-Geisser corrections were used to calculate the degrees of freedom for these values.

The results of the ANOVA for the oxygenation values revealed a significant main effect of Involvement $F(1, 15) = 24.12$, $MSE = 141.113$, $p<.001$, ${}_{p}\eta^2 = .617$., showing that blood oxygenation levels were higher in the active (driving) conditions compared to the passive (replay) conditions. A significant main effect of Task $F(1,15) = 5.07$, $MSE = 21.618$, $p<.05$, ${}_{p}\eta^2 = .253$ demonstrates that blood oxygenation levels were higher in overtaking tasks compared to the following tasks. A significant interaction between Involvement and Task $F(1, 15) = 29.59$, $MSE = 111.887$, $p<.001$, ${}_{p}\eta^2 = .664$ was found and can be seen in Figure 6.4, revealing that the change in oxygenation was significantly higher for overtaking in active condition but no the passive condition.

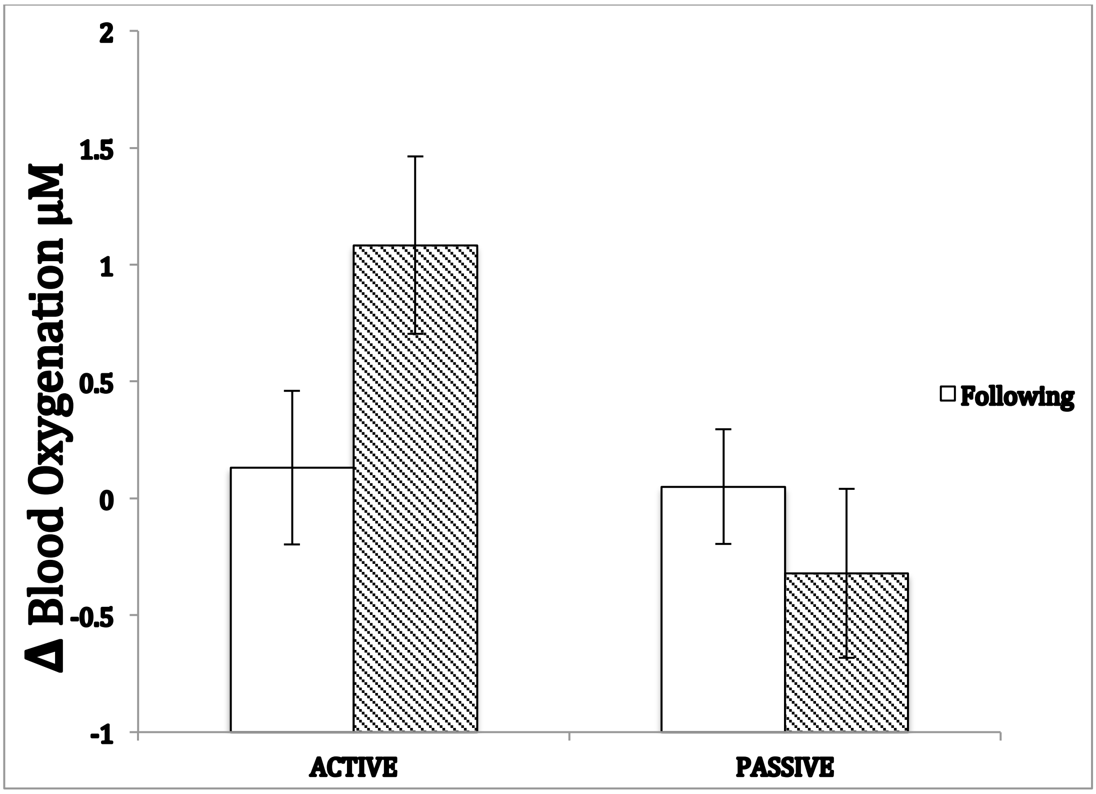


Figure 6.4. Significant interaction between Involvement and Task on the change of blood oxygenation (measured in μ mols) relative to baseline. Error bars represent +/- 1 S.E.

No significant effect of Site [$F(1.765, 26.481) = 0.811$, $MSE = 17.003$, $p=.441$, ${}_{p}\eta^2 = .051$] was found indicating that blood oxygenation levels were roughly equal across all fNIR channels.

6.5.2.2 Hb Levels

For the 2x2x16 ANOVA of the Hb levels, Mauchly's Test of Sphericity was breached for: Site, the Involvement and Site interaction, the Task and Site interaction and the three-way interaction between Involvement, Task and Site. Greenhouse-Geisser corrections were used to calculate the degrees of freedom for these values.

The results of the ANOVA for the Hb levels revealed no significant main effect of Involvement [$F(1,15) = 2.795$, $MSE = 18.312$, $p=.115$, ${}_{p}\eta^2 = .157$], showing that Hb levels were not significantly different in the active and passive conditions. No significant main effect of Task [$F(1,15) = 0.363$, MSE = 1.147, $p=.556$, ${}_{p}\eta^2 = .024$], was found demonstrating that Hb levels were not significantly different in the overtaking and following tasks. No significant interaction between

Involvement and Task was found [$F(1, 15) = 0.204$, $MSE = 0.452$, $p=.658$, ${}_p\eta^2 = .013$]. No significant effect of Site [$F(1.532, 23.025) = 0.861$, $MSE = 14.243$, $p=.409$, ${}_p\eta^2 = .054$] was found, demonstrating that Hb levels were not significantly different across the 16 fNIR channels.

6.5.2.3 HbO2 Levels

For the 2x2x16 ANOVA of HbO_2 values, Mauchly's Test of Sphericity was breached for: Site, the Involvement and Site interaction, the Task and Site interaction and the three-way interaction between Involvement, Task and Site. Greenhouse-Geisser corrections were used to calculate the degrees of freedom for these values.

The results of the ANOVA for the HbO2 levels revealed a significant main effect of Involvement $F(1,15) = 7.140$, $MSE = 27.427$, $p=<.05$, ${}_p\eta^2 = .323$], showing that Hb02 levels were higher in active conditions compared to the passive conditions. No significant main effect of Task [$F(1,15) = 1.495$, MSE = 4.984, p=.240, ${}_p\eta^2 = .091$], was found demonstrating that Hb02 levels were not significantly different in the overtaking and following tasks. A significant interaction between Involvement and Task was found $F(1, 15) = 17.331$, $MSE = 55.555$, $p<.01$, ${}_p\eta^2 = .536$, and is illustrated in Figure 6.5, revealing that the change in HbO_2 was significantly higher for overtaking in the active condition but not the passive condition.

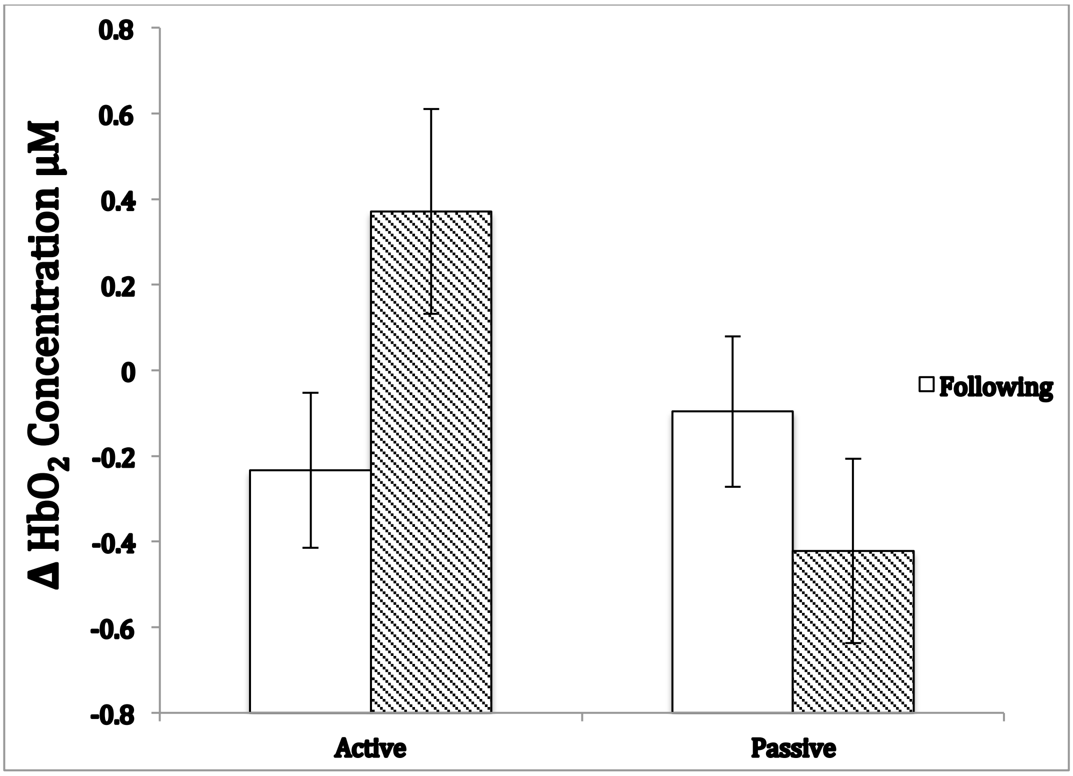


Figure 6.5. Significant interaction between Involvement and Activity on the change in $Hb0_2$ (measured in μ mols) relative to baseline. Error bars represent +/- 1 S.E.

There was no significant effect of Site [$F(4.788, 71.814) = 0.964$, $MSE = 2.116$, $p=.443$, ${}_p\eta^2 = .060$] showing that HbO_2 levels were not significantly different across the 16 fNIR channels.

*Table 6.2. Descriptive statistics, Main Effects and Interactions for each of the fNIR measures used. When P<.05, P = *; when P<.01, P = **; when P<.001, P = ***. Main effect of Involvement, main effect of Task and Involvement*Task interaction values are partial eta squared measures of effect size.*

Involvement Task		Active Follow	 Over.	Passive Follow	 Overt	ME Inv.	ME Task	Task x Activity
Oxygenation								
	mean	0.131	1.083	0.05	-0.32	.617	.253	.664
	sd	1.313	1.52	0.981	1.45	***	*	***
Hb								
	mean	-0.356	-0.465	-0.131	-0.156	.157	.024	.716
	sd	0.918	1.446	0.752	1.137			
HbO_2								
	mean	-0.234	0.371	-0.096	-0.422	.323	.091	.536
	sd	0.727	0.956	0.707	0.862	*		**

6.5.3 Overtaking Event Analysis

In order to investigate whether increased haemodynamic activity during the active overtaking task was primarily due to inhibition or overtaking, an analysis was carried out in which periods during which the participants were overtaking other vehicles was compared with periods in which overtaking was not possible due to the presence of vehicles in the opposite lane. The logic of this analysis was that periods in which the participants were overtaking would be periods during which task workload would be at its highest, whereas the periods in which no overtaking was possible would be periods in which inhibition would be at its highest. Overtaking is likely to result in the highest levels of workload as the participants are required to make the decision to overtake on the basis of the road environment and then use motor skills to complete the overtake without crashing. No overtaking periods are likely to result in the highest levels of inhibition as the participants must inhibit the impulse to attempt an overtake due to it not being safe to do so. The mean change in blood oxygenation during the overtaking and no overtaking periods was selected as the haemodynamic measure as it provides a combined measure of the changes in Hb and HbO_2 throughout these events. Data from all of the participants who successfully completed at least one overtake was included in this analysis. Due to differences in how the participants drove, there were differences in the number of overtakes completed, the time spent overtaking and the time spent where overtakes were not possible, so as much data as possible was included from each participant in the analysis.

Overtaking periods were defined as the point at which the participants fully committed to an overtaking manoeuvre (i.e. the point at which the participants had fully crossed into the next lane to attempt an overtake), until they had re-joined their lane and carried on driving ahead. This process usually took approximately 10 seconds. No overtaking periods were defined as periods of time in which an overtaking manoeuvre was impossible due to there being a vehicle in the opposite lane that was too close to allow for an overtake to be made. Data was taken from periods of at least 10 seconds in which overtaking was not possible in order to provide a more comparable measure with the overtaking periods.

Blood oxygenation was averaged over each of the 16 fNIR channels in order to provide a straightforward and general view of how overall haemodynamic activity changed during the overtaking and no overtaking periods. This decision was also made to help reduce any noise in the data that may have been present due to taking data from a relatively short period of time across all 16 of the fNIR channels.

Table 6.3 lists the mean change in blood oxygenation for overtaking and no overtaking periods for each of the participants included in the analysis. A t-test was then carried out, comparing the mean change in blood oxygenation during overtaking and no overtaking periods averaged across all of the participants. The t-test revealed a significant difference in the mean change in blood oxygenation $t(15)=2.838$, $p<.05$, with significantly higher changes in blood oxygenation during the overtaking periods (M=1.52 S.D.=1.47) than the no overtaking periods (M=1.25 S.D.=1.43).

Table 6.3. Mean change in blood oxygenation (measured in μ mols) for the overtaking and no overtaking periods in the active overtaking task.

	Driving Period	No Overtaking	Overtaking
Oxygenation			
	mean	1.254	1.519
	sd	0.359	0.368

6.5.4 Region Analysis

In order to see if frontal lobe activity showed any specificity to certain areas of the frontal lobes, a regional analysis was carried out by dividing the 16 fNIR channels by their laterality (left or right) and their superiority (top or bottom). A 2x2x2x2 (involvement x task x laterality x superiority) repeated measures ANOVA was carried out on the blood oxygenation values.

The results revealed no significant main effect of laterality [$F(1,15) = 0.004$, $MSE = 0.140$, $p=.949$, ${}_{p}\eta^2 = .0002$], no significant main effect of superiority [$F(1,15) = 0.383$, $MSE = 5.334$, $p=.545$, ${}_{p}\eta^2 = .025$] and no significant interaction between laterality and superiority [$F(1,15) = 2.967$, $MSE = 0.921$, $p=.106$, ${}_{p}\eta^2 = .165$]. There was also no interaction between involvement and laterality [$F(1,15) = 1.041$, $MSE = 0.422$, $p=.324$, ${}_{p}\eta^2 = .064$], involvement and superiority [$F(1,15) = 0.402$, $MSE = 0.341$, $p=.535$, ${}_{p}\eta^2 = .026$], task and laterality [$F(1,15) = 0.665$, $MSE = 0.422$, $p=.428$, ${}_{p}\eta^2 = .0.042$], and task and superiority [$F(1,15) = 0.863$, $MSE =$

1.130, p=.368, $_{p}\eta^{2}$ = .0.054]. These results show that there was no significant difference in blood oxygenation values across the different regions of the frontal lobes that the NIRs sensor was recording from. There were also no differences depending on the task or the involvement.

6.6 Discussion

The results from the NASA TLX workload scores clearly show that perceived workload was higher when the participants were driving in the simulator compared to when they were watching replays. This was to be expected, as the process of driving in the simulator will have inevitably placed more workload on the participants than simply siting in the simulator watching replays. Perceived workload was also higher in the overtaking tasks compared to the following tasks. Compared to the following tasks, the overtaking tasks are likely to elicit more workload as they involve greater task difficulty compared to simply following, as well as greater physical and mental workload that accompany the greater task difficulty. The highest workload scores of all belonged to the active overtaking task, showing that the participants perceived this task to be the most demanding. This is not surprising as this task was designed to be the most demanding task, in terms of task difficulty, as well as physical and mental workload. Overall these results are all in line with the predictions of the study, suggesting that the task elicited the levels of workload they were designed to. The workload scores also mirror the fNIR results on frontal lobe activity during the driving tasks.

The workload scores are similar to the blood oxygenation values, which displayed a similar pattern of results. When looking at the blood oxygenation values, it is clear that there was a higher level of blood oxygenation when the participants were driving compared to when they were watching the video replays. Blood oxygenation was highest in the active overtaking task, just as workload was highest in this task. As this is the condition that required the highest levels of inhibition and workload, it would be predicted that this condition would result in the highest levels of DLPFC activity. This result demonstrates that the DLPFC is most taxed in this condition and that the fNIR device is reliably recording this increase in DLPFC activity. However, unlike the workload levels, blood oxygenation in the passive overtaking task was slightly lower than in the passive following task.

$Hb0_2$ concentrations, whilst not significant for task, were very similar to those of blood oxygenation. $Hb0_2$ concentrations were higher in the driving tasks than the replay tasks and were at their highest levels in the active overtaking task. $Hb0_2$ concentrations were also lower in the passive overtaking task than the passive following task, a result that mirrors the oxygenation findings. Taken together, the oxygenation and $Hb0_2$ concentrations demonstrate increased DLPFC activation during the active tasks, particularly the active overtaking task. The regional analysis carried out with blood oxygenation values failed to find any effect of laterality or superiority however. This demonstrates that the changes in DLPFC activity were largely equal across the areas of the frontal lobe monitored by the fNIR device.

While the Hb results were not significant, the direction of the results does appear logical when examining the complexities of the hemodynamic response. The hemodynamic response is a complicated process with a number of distinct phases. Upon the presentation of a stimulus, some researchers have observed an initial dip in the Blood Oxygenation Level Dependent (BOLD) signal that corresponds to transient increases in the level of Hb (Malonek and Grinvald, 1996). This is thought to reflect a rapid increase in the cerebral metabolic rate of oxygen ($CMRO_2$) before the corresponding increase in cerebral blood flow (CBF). The subsequent increase in CBF is far greater than the increase in $CMRO_2$. This leads to an increase in the levels of HbO_2 and blood oxygenation as well as a reduction in the concentration of Hb, also known as the initial overshoot (Fox and Raichle, 1986). For long stimulus presentations, 20s or more, a plateau effect has been observed in which levels of Hb and HbO_2 remain at a stable level (Bandettini et al. 1992). Inspection of the mean changes in Hb concentration (whilst not significant) reveals that the greatest reductions were in the active tasks, with the single largest reduction in the active overtaking task. This is the opposite pattern observed for the oxygenation and HbO_2 values. As increases in blood oxygenation and HbO_2 concentration will lead to the reduction of Hb concentration, the reductions in Hb do appear logical. Tsunashima and Yanagisawa (2009) observed this effect when using fNIR in their driving study.

The fact that the active overtaking task produced the largest increases in oxygenation and HbO_2 is potentially an indication of the DLPFC being utilised for inhibition. The overtaking task was chosen as previous research (e.g. Clarke et al. 1998) has reported that inhibition is of key importance when attempting an overtaking manoeuvre and inhibition is thought to be heavily dependent on areas of the frontal lobes, including the DLPFC (Alvarez and Emory, 2006). Although this study is not able to fully disentangle the relative effects of workload and inhibition on the DLPFC activation, the oxygenation and HbO_2 values in the active overtaking task may be too large to be explained purely as a result of increased workload. For instance, the difference in oxygenation and HbO_2 between the active and passive overtaking tasks are much larger than the differences between the active and passive following tasks. The difference between the active and passive following was not nearly as large as the difference between the active and passive overtaking tasks. It is possible that the extra difference in oxygenation and HbO_2 changes between the overtaking tasks results from the inhibition required in the active overtaking task.

The data from the overtaking and no overtaking periods may provide some indication as to the relative effects of workload and inhibition on DLPFC activity in this study. If the increase in DLPFC activity is primarily due to inhibition we could expect that the highest levels of blood oxygenation to correspond with periods in which the desire to overtake was being inhibited (i.e. during the no overtaking periods). If the increase in DLPFC is primarily due to increased workload we could expect that the highest levels of blood oxygenation correspond with periods in which overtaking was occurring. The results revealed that the highest levels of blood oxygenation occurred during the periods in which the participants were overtaking other vehicles. This supports the hypothesis that increases in DLPFC activity is primarily driven by task workload. However, the data also reveals that the change in blood oxygenation increased during the periods in which no overtaking was possible, suggesting that task inhibition may also contribute to increased DLPFC activity.

Overall, the main driver of differences in DLPFC activity appears to be from the differences in task workload, although task inhibition may also play a role. This is in line with previous research has investigated the role of the DLPFC in driving with the use of fNIR. Tsunashima and Yanagisawa (2009) reported that increased DLPFC activity during driving was associated with increased task workload. The subjective workload scores in this study are also in line with those of Tsunashima and Yanagisawa (2009), with higher scores being associated with the higher workload conditions.

Kojima, Tsunashima and Shiozawa (2006) used fNIR to measure DLPFC activity while participants drove in a train simulator. Their findings indicated that increased DLPFC activity was associated with increased task difficulty (i.e. braking to stop at a train station compared to driving at a constant speed along the track), which will have resulted in increased task workload. This effect was especially apparent in the untrained driver who produced larger levels of DLPFC activity and the highest recorded level of activity when they failed to properly stop at one of the stations. These findings suggest that increased task workload, stemming from increased task difficulty, result in increased DLPFC activity. This is again in line with the findings of this current study. The following task in study 4 is comparable to the part of the Kojima et al. (2006) study in which the participants drove at a constant speed from one station to the next. Meanwhile the overtaking task in study 4 can be compared to the process of stopping the train at one of the train stations. Both tasks require skilled manipulation of the driving controls and the judgment of the driver to complete the manoeuvre successfully. In both studies the tasks that required skilled manoeuvring resulted in significantly greater DLPFC activity than the tasks that involved following the road (or track) ahead.

The results in this study support the research carried out by Yoshino et al. (2013). The constant driving velocity task in their experiment is similar to the following task in this study. In their study, driving at a constant velocity resulted in the lowest levels of prefrontal cortex activity. They put this down to the reduced workload of the task, namely from the limited need to make adjustments to the driving controls in the constant velocity task. A similar argument can be made for the active following task in this study, whereby the lower levels of frontal lobe activity compared to the overtaking task were most likely due to differences in workload between the tasks. Ultimately the amounts of driving inputs in the active following task are not as large as those in the active overtaking task. In the study by Kojima et al. (2006) the lowest levels of frontal lobe activity were recorded when the participants were driving at a constant velocity between stations, a task that required minimal control inputs. In all of these studies, the tasks with minimal driving inputs and lower levels of workload produced minimal activity in the frontal lobes. Meanwhile tasks that required driving inputs in order to change speed or direction resulted in significant levels of DLPFC activity. This lends support to the view that workload is driving activity in the frontal lobes while driving.

Yoshino et al. (2013) found during the acceleration task activity in the DLPFC increased significantly. They posited that the DLPFC was involved in internal monitoring during the driving task in order to make sure that the target speed was achieved within the allowed distance. One of the functions of the DLPFC is to control attention while executive functions are being executed (Shallice, 1982). It is possible that the DLPFC was engaging in such a process in the active overtaking task. When executing a successful overtake in the overtaking task it is vital to accurately judge both the speed and distance required to complete the manoeuvre without having an accident. This could involve determining if there is enough space in front to overtake, accelerating at the right rate to get past a vehicle, and then decelerating once you have retuned to your own lane. All of this has to be achieved while avoiding any oncoming vehicles.

A major point brought up in the research by Yoshino et al. (2013) is the importance of visual processing in the frontal eye field. They found that the frontal eye field was significantly activated during all driving tasks. However, the level of activation was dependent on the driving task. Driving at a constant velocity produced lower levels of activation than tasks that involved deceleration and acceleration. The amount of activity in the frontal eye field also varied depending on the rate of change in speed, with greater deceleration or acceleration leading to greater activation. Yoshino et al. (2013) theorised that this was due to increased voluntary eye movements in three-dimensional space. The frontal eye field is responsible for two types of voluntary eye movement: vergence eye movements responsible for the perception of depth and pursuit eye movements from side-to-side (Gamlin and Yoon, 2000). Fukushima et al. (2002) reported that one of the most important functions of the frontal eye field is the formation of eye movements in three-dimensional space. Increased frontal lobe activity in the active overtaking task may partly have been due to complex visual processing for overtakes leading to increased activity in the frontal eye field.

Overall the results of this study suggest that fNIR is able to reliably record activity in the DLPFC, as the highest levels of DLPFC activity occurred in the active overtaking task (which has the highest levels of both workload and inhibition). The finding that frontal lobe activity was higher in the active tasks compared to the passive watching tasks indicates that fNIR is not simply recording brain activity associated with visual stimulus. The fact that the fNIR data and workload data are similar further indicates that fNIR is reliably recording activity related to workload. These results are also in line with those of Tsunashima and Yanagisawa (2009), who found a strong correlation between frontal lobe recordings using fNIR and subjective workload ratings using the NASA TLX. The results of this study further suggest that the NASA TLX is a reliable measure of perceived workload. The pattern of blood oxygenation data suggests that increased frontal lobe activity is related to increased task workload. This is supported by previous research that has recorded increased DLPFC activity related to workload during simulated driving tasks. Increased task inhibition may also play a role in the levels of DLPFC activation during the tasks.

However the design of this study does not allow for a specific conclusion on this subject, as the relative effects of workload and inhibition on DLPFC activity are difficult to disentangle. The large differences in frontal lobe activity between the active overtaking and active following tasks indicate that the differences in task workload were too large. As a result, any differences in inhibition required in the tasks would have been extremely difficult to disentangle. This issue is compounded by the fact that subjective score for task inhibition were not collected. The possible role of inhibition in DLPFC activity during driving does warrant further investigation.

There are several improvements that could be made to this study. One of the limitations to this study related the difficulties in synchronising the fNIR data to events of interest such as overtakes. There were also difficulties in identifying comparable non-overtaking moments and having overtakes that could be compared between the participants (i.e. some people did not complete any overtakes and other differed in terms of the number of overtakes completed). One potential improvement would have been to come up with a system of synchronising the overtaking manoeuvres in the passive overtaking task with the fNIR recordings. This would have made it possible to see if there were any changes in haemodynamic activity associated with the overtaking manoeuvres taking place in the replays. A possible solution to these issues would be to strictly control the behaviour of the AI vehicles so that there was a guarantee for at least one overtaking chance to occur during the active overtaking tasks. The timings of this event could also be controlled, which would also allow for easier synchronisation with the fNIR recordings. This solution could also be used to guarantee that at certain times during the active overtaking condition, there was absolutely no chance of overtaking. This would make it much easier to tie events of interest (i.e. overtaking and no overtaking possible periods) to any corresponding haemodynamic changes.

Chapter 7- The effect of Age, Driving Experience and Gender on Driving and DLPFC Activity

7.1 Abstract

Driving research has repeatedly demonstrated that the age, experience and gender of a driver has an impact on their likelihood of being involved in a collision, with young, inexperienced and male drivers being significantly overrepresented in crash and accident statistics. Therefore, it is important to establish what underlying factors contribute to this phenomenon in order to improve safety for all road users. One possible explanation for the high number of traffic accidents that young and inexperienced drivers are involved in is a lack of maturation in the frontal lobes and associated cognitive functions. In order to investigate this, a simulated driving study was carried out using fNIR to record DLPFC activity during a number of simulated overtaking tasks (in which the traffic density was manipulated). Male and female participants with a range of ages and levels of driving experience were recruited in order to explore any possible effects age and driving experienced may have on DLPFC activity during driving. The results revealed that DLPFC activity was higher when overtaking than following. However, there were no observed effects of age, driving experience or traffic density on frontal lobe activity.

7.2 Introduction

The results of study 4 revealed that fNIR was able to reliably record activity in the DLPFC during simulated driving. The results of the simulated overtaking study suggest that differences in frontal lobe activity are largely due to differences in task workload. There was also some suggestion that inhibition may have been responsible for some of the changes in DLPFC activation. However the design of the study limited the ability to draw a firm conclusion on this subject. The next step is to continue investigating the role of the DLPFC in driving with some methodological changes. Furthering the need to continue investigating this subject is the fact that a number of other factors may impact on the role of the frontal lobes in driving. One factor that consistently emerges in driving research is the effect that age and experience have on driving. Young and inexperienced drivers are hugely overrepresented in crash and fatality statistics. Research has shown that the frontal lobes and the functions related to them do not reach full maturation until about 25 years of age. This is eight years after the legal driving age in the UK. As a result it is possible that the lack of maturation in the frontal lobes (and the cognitive functions associated with them) is one of the reasons that young novice drivers are involved in so many crashes. Research has also shown that gender can have a significant effect on driving behaviours. The use if NIRs may allow further insights into these factors.

7.2.1 The Effect of Age and Experience on Driving

As mentioned in Chapter 1, a consistent finding in driving research is the significant over-representation of young novice drivers in crash and fatality statistics. An OECD report (ECMT, 2006) revealed that globally, 16-24 year old drivers are significantly over-represented in crash and fatality statistics. Despite the age group representing only 10% of the populations of OECD countries, young drivers represent 27% of all drivers killed in OECD countries. The report also found that the death rate for 18-24 year olds is about double that of older drivers and that crashes are the single greatest killer of 15-24 year olds in OECD countries (ECMT, 2006). The report detailed that the fatality rate per 1 million drivers in each age group dropped significantly from approximately 90 for 18-24 year olds to approximately 40 for 25-64 year olds. Despite the fact that overall fatality rates have fallen over time, due to improvements in road and vehicle safety, the large difference in fatality between young drivers and other age groups remains a factor across OECD countries.

Statistics released by the Australian Government in 1996 showed that serious crash involvement amongst those aged 17 - 20 years was over 12 times that of drivers aged 45 - 49 years, while young drivers are involved in significantly more crashes than all other driving groups except for the elderly (Wylie 1996). Similar statistics and findings have been released in the United States (Levy, 1990) and in New Zealand (NZ Ministry of Transport, 2010). Figures from the UK reveal that male drivers between the ages of 17 and 20 are involved in an average of 440 injury accidents per 100 million km driven. This is compared to an average for all male drivers of 106 accidents per 100 million km driven (Forsyth, 1992). This age bracket appears to be especially prone to accidents as there was a drop off to 180 injuries per 100 million km driven for 20-24 year old males, all though this is still significantly higher than the average for all males (Forsyth, 1992). Findings from the UK Select Committee on Environment, Transport and Regional Affairs 19th report (1999) detailed that although they only hold 11% of total driving licenses, 17-24 year olds are involved in 25% of fatal or serious accidents each year. Male drivers aged 17-24 also had fatality rates 10 times higher than those of males aged 35-54. More recent data from the UK reveals that the likelihood of being killed or seriously injured on the road is highest for those in their late teens and early twenties, with 18 year-olds have a casualty rate 8 times that of 68 year-olds (Chowdhury and Kilbey, 2010).

By nature, young drivers are inexperienced drivers and inexperience has been found to be a contributing factor in the accident rates of young drivers. This is unsurprising given the fact that inexperienced drivers will lack experience with vehicle control, driving in certain weather conditions and dealing with certain driving hazards for example. In an analysis of crash statistics in Michigan, Waller et al. (2001) reported that crash rates reduced by 17% for each year of driving. This increased to 22% after the first year of driving. A 41% drop in crash rates across the first 7 months of driving per 10,000 novice drivers was also found, indicating that the effect of experience is perhaps most pronounced during the first few months of driving. There was also an affect of age, with greater reductions at younger ages. There was a 56% fall for 16 year olds and a 30% drop for 17 year olds. This indicates that experience an age interact with regards to crash rates. McKnight and McKnight (2003) reported that the effect of inexperience was much greater than that of age for young novice drivers. This is supported by research from Wells et al. (2008) who reported that experience had a greater effect on accident liability than age, with drivers of all ages at particular risk in the first 6 months of qualified driving. Maycock, Lockwood and Lester (1991) investigated the effects of age and experience on accident liability. They carried out a study on newly qualified British drivers of different ages (17, 20, 25, 36 and 50) who travelled approximately 12,000 km/year. They reported that crash risk during the few years of solo driving decreased by 31% due to age and 59% due to experience. This finding demonstrates the importance of both age and experience in the accident liability of drivers.

Further evidence of the effect that age and experience have on accident liability comes from Forsyth et al. (1995) who reported that accident rates for young drivers decrease rapidly as a function of both age and experience. Harrington and McBride (1970) reported that the frequency of each type of traffic violation decreased with age. Mayhew, Simpson and Pak (2003) found that during the first few months of driving, 16 year olds are involved in more crashes than recently licensed older drivers. McDonald (1994) argued that experience tended to have a greater effect on older drivers as well as female drivers, but for young males, age seems to be the main factor in the reduction of accident liability. Lewis-Evans and Lukkien (2007) reported that crash rates are highest for novice drivers during the first month with a dramatic decrease in accident liability over the next few months. McCartt et al. (2005) argued that the younger the driver, the more age contributes to crashes than inexperience does.

This research makes it clear that both driver age and experience play a significant role in the crash liability of young novice driver, although the exact contribution of each appears to vary depending on certain circumstances. Jonah (1986) even stated 'the attempt to separate the two concepts may well prove fruitless'.

Research has also found that there are differences in the nature of accidents that young novice drivers have. Maycock (2002) compared the accident characteristics of young novice drivers and older drivers. Maycock reported that young drivers are largely involved in particular accident types: single vehicle accidents, accidents at night, accidents at bends, accidents while under the influence of alcohol, speed related accidents and overtaking accidents. Young drivers were also found to have a greater number of accidents in the early morning and evening. These accidents tended to be single car accidents. In support of these findings, the incidence of rear end collisions has been found to be particularly high amongst younger drivers and only begins to decrease after the age of 26 (Yan, Radwan & Abdel-Aty, 2005). Williams, Ferguson and Shope (2002) reported that age is a significant factor in the seriousness of crashes, with young drivers more likely to be involved in serious crashes than older drivers.

Clarke, Ward and Jones (1998) investigated the effect of age on the type of car accidents. They found that 'return-and-lose-control' incidents (where the driver loses control of the vehicle after returning to the nearside lane after completing the overtake) were heavily associated with young drivers (16 to 22 years of age). The authors also reported on different causes for accidents between young novice and older drivers. They found that the most common form of accident as a result of an overtaking manoeuvre was a collision with a vehicle turning right. For young novice drivers these accidents were the result of making an incorrect overtaking decision, but for older drivers they were due to them making a faulty decision to right turn. Clarke, Ward and Truman (2002) reported that young novice drivers tend to have specific problems on the roads. This includes having accidents while driving at night, speeding, having single vehicle accidents at bends, rear end collisions and having accidents when making right turns.

7.2.2 Differences Between Young Novice Drivers and Older Experienced Drivers

Research has indicated that there are a number of contributing factors that lead to the extremely high crash rates and different crash types of young novice drivers including: poor risk perception, inferior hazard perception, impulsivity, thrill-seeking, sensation-seeking, and risky driving behaviours (e.g. speeding), a propensity to take risks, poor vehicle control, overestimating driving skills, underestimating the task demand of driving and sensitivity to peer influences in adopting inappropriate norms (Williams, 1998; Lee, 2007). These factors have been categorised into the young driver problem “due to driving inexperience” or the young driver problem “due to age related risk taking” (Senserrick, 2006).

Bragg and Finn (1982) carried out a comprehensive study on the effect of age on driver’s perception of accident risk. The attitudes of young drivers (18-24 years) and older drivers (38-50) were compared. Young drivers differed in their perception of risk, viewing some situations as more risky and others less risky than the older drivers. The overall risk of having an accident was rated as being higher in young drivers, however they viewed certain driving behaviours such as speeding and tailgating as being less risky. The young drivers were also more likely than older drivers to view their peers as being more at risk than they themselves were. In effect, the young drivers perceived themselves as being different from their peers (namely being better drivers and at less risk than their peers). The authors suggested that the lower perception of risk in young drivers might have been due to them having a greater level of confidence in their skill as drivers.

Following on from the work of Bragg and Finn (1982), Matthews and Moran (1986) compared young (18-25) and older (35-50) male drivers in terms of their perception of driving risk and how confident they were in their driving ability. As Bragg and Finn (1982) had predicted, the authors found that young drivers do overestimate their own driving abilities as well underestimating their risk of being involved in an accident. These factors were interrelated. The higher a driver estimates their own ability the lower they estimate the level of risk. The young drivers ultimately viewed themselves as having the driving ability of the older more experienced drivers. They also viewed driving hazards as being a risk to their peers but not themselves, as well as perceiving their driving abilities to be superior to their peers (something which older drivers did not do). The research showed that young drivers displayed a disassociation between perceived and actual driving ability, and viewed themselves as being immune to certain driving risks even though they viewed their own peers as being liable to these risks. Rhodes and Pivik (2010) reported similar results, with young drivers having a lower perception of risk as well as showing more enjoyment for risky behaviours. This was especially prevalent in young male drivers. The practical effects of this risk taking behaviour were demonstrated Clarke, Ward and Truman (2005) who found that the majority of young driver accidents were caused by risk taking, as opposed to deficits in driving skill.

Research has also focused on whether there are any underlying cognitive differences in young novice drivers that may explain why they are more dangerous and accident-prone. Jongen et al. (2011) studied the cognitive mechanisms that underlie the risk taking behaviour in young novice drivers. They hypothesised that risky driving was the result of a lower capacity for cognitive control (response inhibition) and the potential of reward. Two groups of young and inexperienced drivers (between 17-18 or 22-24 years of age) were tested. A Stop Signal Reaction Time (SSRT) task was used to measure inhibitory control and a driving simulator task was used to determine risky driving behaviour (e.g. collisions, speeding, standard deviation of lateral lane position (SDLP)). Reward was manipulated by instructing the participants to complete their second run of the driving simulation as quickly as possible. The faster they finished the better chance they had of receiving a monetary reward. However, each driving violation or collision they were involved in would result in a one-minute penalty. This did not apply to speeding though.

The results of the study showed that inhibitory control was lower in the 17-18 age group than in the 22-24 age group, demonstrating that inhibitory control continued to develop even at this age. Drivers with lower inhibitory control demonstrated more risky driving with a higher SDLP. Risky driving was predicted by reward, with more dangerous driving occurring when drivers were offered rewards for completing the simulated drive as quickly as possible. This study indicates that inhibition is developing in early adulthood, and as a result young drivers lack the inhibitory capabilities of older drivers. This may be one of the factors that cause the differences between young drivers and older drivers.

7.2.3 Cognitive Development in Adolescence and Young Adulthood

The research carried out on young drivers makes it clear that they are far more likely to take risks, perceive behaviours as being less risky and overestimate their driving ability. These differences may relate to the maturation of the frontal lobes and frontal cognitive functions (e.g. ability to inhibit risky behaviours, impulse control). This may be an important factor in the differences between young novice and older drivers. As mentioned in Chapter 6, low inhibition and high impulsiveness are associated with dangerous and risky driving. This fits with research that has shown that during adolescence, frontal cognitive functions are still developing and show marked difference to those of adults. These factors may be associated with the dangerous driving of young novices. It is possible that underdeveloped cognitive functions such as these may be an underlying cause for dangerous driving of young novices.

As Chapter 4 explained, a key role of the frontal lobes is to exert 'executive control' over the brain. This includes: impulse control, planning, allocating attention, decision-making and judging risks (Weinberger et al. 2005). While other areas of the brain and their functions have fully developed, frontal functions are still developing between adolescents and adulthood and result in a period of heightened vulnerability through risk taking (Steinberg 2007, 2008). This period is marked by deficits in frontal cognitive functions accompanied by increased risk taking and stimulus seeking (Weinberger et al. 2005), which may in turn be one of the causes of the dangerous and accident prone driving of young drivers.

Steinberg (2007, 2008) proposed that this increased risk taking during adolescence is the result of significant and sudden increases in dopaminergic activity within the socio-emotional system during puberty that leads to increased reward seeking. The issue is that this increase in reward seeking takes place before the frontal system that mediates cognitive control has developed. This system gradually develops over adolescence to provide more impulse control. The temporal gap in the maturation of these systems results in a period of increased risk taking. Steinberg proposed that sensation seeking and impulsivity might provide the mechanism for this increased risk taking. This increase in risk taking behaviour can easily translate into dangerous driving (e.g. speeding, drink-driving) which is associated with crashes.

Research has shown that performance on tasks that test executive functions, as well as brain activity associated with executive function shows significant changes from adolescence through to adulthood. A study by Luna et al. (2001) revealed that adolescents produced greater amount of activation in the DLPFC than children or adults during a response inhibition task. These findings suggested that the adolescents were more reliant on the frontal executive network to produce response inhibition. In other words, the adolescents had to evoke a greater response from the DLPFC to produce similar levels of response inhibition, suggesting inefficiency in these functions. Yurgelun-Todd (2007) found that activity in the PFC resulting from inhibitory tasks (such as the Stroop task) become more localised with age, while diffuse activity reduced. Overall, research has shown that cognitive development through adolescence is associated with increasing efficiency in executive control (Rubia et al. 2000), while frontal activity becomes more focal and specialised (Durston et al. 2006). This development appears to lead to the most important change in adolescent cognition, which is the development of a fully functioning executive control system (Keating, 2004; Donald, 2001). As this system is still developing through adolescence and into early adulthood, it may help explain the cognitive mechanisms that underlie the driving behaviours of young drivers.

Dahl (2001) has described the delay in frontal lobe maturation as being like starting the engine without a skilled driver behind the wheel. With regards to driving, a lack of fully developed cognitive functions may result in young drivers not having the necessary cognitive functions to properly perceive risk, inhibit dangerous responses etc. To make matters worse they are predisposed to risky, impulsive and sensation seeking behaviour. This very much indicates that fully developed executive functions are vital for driving.

7.2.4 Frontal Lobe Maturation

As the above research demonstrates, executive functions continue to develop through adolescents and early adulthood. An underlying factor in this phenomenon may relate to the physiology of the frontal lobes. A significant amount of research has shown that there are corresponding maturational changes taking place in the frontal lobes as executive functions develop. It is possible that the stage of development of the frontal lobes plays an important role in the propensity of young novice drivers to be involved in road accidents. For example, a lack of maturation in frontal regions may lead to a lack of fully developed frontal lobe cognitive abilities. This in turn may lead to the dangerous driving and increased crash propensity of young drivers.

Up until recently, the prevailing thought was that the brain was fully developed by the end of childhood or early into adolescence. However, more recent neuroscience research has shown that this is not true and the development of the brain follows a more complex timeline. Sowell et al. (1999) used MRI to study differences in the brains of adolescents (12-16 years) and adults (23-30 years). Their study revealed that while there were little differences in the parietal, temporal and occipital lobes there were large differences in dorsal, medial and lateral regions of the frontal lobes. Their research revealed reductions in grey matter in frontal regions between adolescence and adulthood. This is most likely a result of increased myelination in peripheral regions of the cortex. Myelination results in an increase in the speed at which impulses propagate along neurons. A result of this increased myelination may be enhanced cognitive processing in adulthood due to faster and more complex information sharing in the brain (Weinberger et al. 2005). This may in turn lead to the ability to carry out complex tasks such as driving, which requires the utilisation of multiple skills and regions of the brain simultaneously. The dopamine inputs in the PFC are also still developing during adolescents providing another mechanism through which impulse control and improved cognitive processing develop Weinberger et al. (2005). Overall, frontal lobe maturation was not complete until at least 25 years of age. This is a full 8 years after the legal driving age in the UK (and similarly long for most other countries) as well as being the age at which fatality rates for drivers fall significantly compared to those between the ages of 18 and 24 (ECMT, 2006).

The MRI study also revealed striatal changes, namely in the globus pallidus and the putamen, again the result of reductions of grey matter caused by increased myelination. The frontal cortex mediates striatal motor functions and the striatal structures are themselves involved in cognitive functions such as learning, which is linked to frontal lobe functions (Rolls, 1994) and improves throughout adolescence (Levin, Culhane and Hartmann, 1991). The authors argued that this was an indication of temporal and functional relationships between the simultaneous post adolescent reductions in grey matter density in frontal and striatal regions. In support of these findings, Yakolev and Lecours (1967) carried out post-mortem studies and found that myelination in frontal and parietal areas progressed over a protracted period of time. This myelination continued long into the third decade of life.

Sowell et al. (2001) used Voxel Based Morphometry combined with cortical-matching techniques to study tissue changes in the developing brain. They found that between adolescence and adulthood, there were significant reductions in grey matter density in the frontal lobes. Although this effect was also present between childhood and adolescence, it accelerated significantly between adolescence and adulthood. This was especially prevalent in dorsal frontal areas. In the same dorsal areas that reductions in grey matter were found, the authors found localised areas of growth between adolescents and adults. Localised growth was also found in lateral and orbital frontal cortex. This was accompanied by diffuse shrinkage in frontal areas that surrounded the areas of growth. The authors also reported that the areas of grey matter reduction corresponded almost exactly with areas that had undergone increases in size. They attributed these events to post-adolescent maturation.

Lebel et al. (2008) used diffusion tensor imaging to study the brains of people between the ages of 5 and 30 years. They found that major white matter tracts, subcortical white matter and deep grey matter changed with age. There was a considerable difference in these changes in different regions of the brain. Areas with fronto-temporal connections were found to develop more slowly than others. The cingulum and uncinate fasiculus, both of which are important fronto-temporal connections, were found to reach 90% development only after 25 years of age. Like Sowell et al. (1999) found that the globus pallidus and putamen changed into adolescence, Lebel et al. (2008) found that these areas only mature after 25 years of age. Other studies utilising DTI have also found that frontal areas mature later than other areas of the brain (McKinstry et al. 2002).

The research on the maturation of the frontal lobes shows that frontal areas continue to develop beyond adolescence and into early adulthood. This maturation appears to be complete around the age of 25. Meanwhile cognitive functions associated with the frontal lobes appear to maturate on a similar timescale. Combined with the finding that the factor of age disappears as a factor for car crashes (Mayhew, Simpson and Pak 2003) supports the idea that maturation of the frontal lobes may an important factor in the effect that age has on driving.

There has been some research that has investigated the effect that driving age and experience have on frontal lobe activity. Harada, Nashihara, Morozumi, Ota and Hatakeyama (2007) investigated the effects of age and experience on frontal lobe activity during driving. They compared cerebral activity in the prefrontal cortex between young adults (who were either experienced or inexperienced at driving) and elderly people while they drove in a simulator. The participants were either young adults (who were in their early twenties) or elderly drivers (who were in their sixties). The experienced young drivers drove every day whereas inexperienced drivers had not driven since they received their driving licenses. Activity in the PFC (consisting of Hb and HbO_2 levels; and total Hb concentrations) was measured using fNIR and time-resolved spectroscopy (TRS).

They found that inexperienced young adults produced greater frontal lobe activity during driving than their experienced counterparts. Less experienced drivers also failed to look at their mirrors as often as the experienced young adults. This pattern of behaviour regarding mirror checks was also found in the elderly male drivers. However the frontal lobe activity of the elderly male drivers was lower. The less experienced young adults also produced significant increases in frontal lobe activity when braking to stop at traffic lights. This effect was not found in the experienced young adults and elderly drivers. Prefrontal activity at rest was found to lower in the elderly participants than the young adults. There was also less variation in frontal lobe activity during driving from the elderly participants. The greatest level of variation in frontal lobe activity during driving came from the inexperienced young adults. The authors concluded the inexperienced young drivers were less adapted to driving than the more experienced and elderly drivers.

7.2.5 The Effect of Gender on Driving

Another factor that continually arises in driving research is that the driving behaviours and characteristics of males and females are different. Young male drivers aged 15 -24 are at higher risk than female drivers of the same age group or older male drivers (Williamson, 2003). In fact, males have higher fatality rates than females across all age groups, with the most striking difference in the late teens and early twenties (ECMT, 2006). A 2007 report by the World Health Organisation revealed that road traffic accidents were the leading cause of death among young people between ages 10 - 24 years, with young males at higher risk for road traffic fatalities than females in all age groups under 25. 17-20 year old males in the UK were involved in 440 accidents per 100 million km whereas women of the same age were involved in 240 accidents per 100 million km (Forsyth, 1992). Mills et al. (2008) reported that male respondents reported more accidents than female respondents at 6, 12, 24 and 36 months after they passed their driving tests. Research has also shown that males involved in fatal car crashes are more likely to have been speeding than females, while the fatality rate for males is more than 3 times that of females (NHTSA, 2009). Harre et al. (1996) reported that male drivers report engaging in more risk taking behaviours such as speeding and drink driving than females of the same age.

There also appear to be differences in terms of how age and experience influence male and female drivers. As mentioned previously, MacDonald (1994) found that age appeared to be the main factor for young males in relation to crash rates, while for older drivers and young females experience had the greatest effect. Meanwhile Yannis (2007) found that regardless of age, male drivers were more likely to be involved in traffic accidents than female drivers. The most prominent difference was found in 18-20 year old drivers, where males were 12 times more likely to be involved in traffic accidents than females. Meanwhile Wells et al. (2008) reported that accident liability changed more rapidly with the age of passing the test for male drivers than for female drivers.

A number of reasons have been put forward to explain the differences between male and female drivers. Lajunen and Parker (2001) found that driver anger and aggression in males was negatively correlated with age, with anger and aggression reducing with age. Meanwhile, driver aggression in females was negatively correlated with annual mileage. These findings indicate that the effects of age and experience on driving are different for males and females, with males becoming less aggressive with age and females becoming less aggressive with experience. Begg and Langley (2001) investigated the changes in the prevalence of risky driving behavior in young drivers when they were 21 and 26. They found that risky driving was more common in male drivers but by 26 years of age most had "matured out" of risky driving behaviours. The fact that young male drivers are more likely to take risks may explain why they are involved in more crashes. It also suggests that there may be specific characteristics in young male drivers that explain their driving behaviour.

Rhodes and Pivik (2010) investigated the relationship between risk perception, positive affect and risky driving. They reported that male drivers were more likely to engage in risky driving behaviours. Interestingly, they found that this risky driving behaviour was best predicted by low risk perception and positive affect (i.e. male drivers reported enjoying risky driving). This research suggests that male driving behaviours may be in part caused by low perception of risks and enjoying risky behaviours. This is supported by research from Yagil (1998) who reported that males were more selective in obeying traffic laws and that this was associated with males perceiving traffic violations as being less dangerous than females

Lonczak, Neighbors and Donovan (2006) investigated the causes of gender differences in driving. They found that men reported committing more traffic violations and receiving more driving related injuries. Men also demonstrated a greater association between the number of traffic violations committed and sensation seeking. This fits with research from Ulleberg (2001) who did an analysis to identify risky driver groups among adolescents. The author found that males with high levels of sensation seeking constituted one such group. Lonczak et al. (2006) also reported that although levels of driving anger were no different in males and females, the way this anger was expressed was different.

Research has also investigated the different driving styles and personality traits that men and women have, and if this contributes to accident involvement. Holland, Geraghty and Shah (2010) investigated how the locus of control (LOC) interacts with gender and driving experience to affect driving style. Female drivers were found to be more dissociative, anxious and patient drivers. Male drivers scored higher on angry, risky and high-velocity driving styles. The authors argued that these driving styles might be major contributing factors in the driving accidents that men and women have. With regards to LOC, females demonstrated a more external LOC than males. The external LOC was correlated with the driving styles of women (e.g. anxious). There was also an interaction with driving experience, with increasing driving experience providing a positive effect for men (reducing angry and high-velocity, increasing carefulness) and a negative effect for women (increasing angry and higher velocity, reducing carefulness). The authors argued that the results demonstrated that a high internal LOC has negative effects for young male drivers. This may be related to the fact that young drivers have been found to overestimate their own driving ability (Matthews and Moran (1986), with driving ability being an internal LOC.

This research demonstrates that there are significant differences in the driving behaviours of males and females. There also appear to be interactions between gender and age with young males displaying particularly risky driving behaviours and being the most likely to be involved in crashes on the road.

7.3 Study 5

In order to assess whether or not frontal lobe maturation has an effect on driving it will be necessary to recruit a representative sample of drivers to participate. This sample of drivers will need to consist of drivers who differ in terms of age, driving experience and gender. By comparing fNIR data between these different groups of drivers it may be possible to establish if there are differences in DLPFC activity as a function of the driver's age, experience and gender. At the same time, the role of workload and inhibition in DLPFC activity during driving can be further investigated.

In order to further the research of Study 4, this study aims to investigate the role that the DLPFC plays in driving as well as examining what effect age, driving experience and gender have. In this experiment, overtaking will again be compared with following in order to manipulate the levels of workload and inhibition required. However this time there will be four overtaking tasks that differ in levels of traffic density. The idea here is that the differing levels of traffic density will provide incremental changes in workload and inhibition. The purpose of this is to avoid the large differences in task workload and inhibition that were present in Study 4 between the active overtaking and following tasks, and between the active and passive tasks. However, the aim is that the changes in inhibition will be greater across the densities as the primary effect of increasing the traffic density is to make overtaking more difficult. As a result of this the participants may be forced to spend more time inhibiting unsafe manoeuvres, as the increased number of cars on the road will make safe overtaking opportunities scarcer. This potentially allows for a clearer idea of what role workload and inhibition play in DLPFC activation. Given that young novice drivers seem to struggle with inhibition and impulse control, an overtaking task appears to be a suitable way of testing the effects of age and experience of DLPFC activity. The passive tasks used in Study 4 will not be used in this study. A modified version of the NASA TLX workload scale than contains a measure of inhibitory control will be used. This measure has been included in order to provide a subjective measure of inhibition to compare with the subjective measures of workload.

It is predicted that frontal lobe activity, subjective workload and subjective inhibition scores will be higher in the overtaking tasks than the following task. These values will also increase as the traffic densities increase. Differences in DLPFC activity, subjective workload and subjective inhibition scores are expected as a result of differences in age, experience and gender. It is also predicted that the number of overtakes completed will fall with increasing traffic density.

7.4 Methodology

7.4.1 Participants

A power analysis revealed that for a medium effect size (F = 0.25), a total of 32 participants (8 in each of the between-subject groups) would be required for this study.

32 participants (16 male and 16 female) were recruited to take part in the study. All of the participants were first or third year undergraduate students from the University of Nottingham who were fully qualified drivers. All of the participants had normal or corrected-to-normal vision. In order to explore the impact of age and experience, participants were selected according to their driving experience (determined by the total number of miles they had driven since gaining

their full driving licence) and age. Participants who had driven 10,000 miles or more were placed into the ‘experienced’ group, while those who had driven less than 5,000 miles were placed into the ‘novice’ group. This split was chosen as it provided as large as possible a difference in the maximum and minimum number if miles driven while allowing the between subject groups to remain equal in size. This split also allowed for a large difference in the mean number of miles driven by the experienced and novice drivers.

The total number of miles driven was chosen as the measure of driving experience rather than the number of years driving, which is often used as a measure of driving experience, because the range of the participant’s experience in terms of years driving was limited (due to the sample of participants) whereas there was a large range in the total number of miles driven by the available participants. Defining driving experience on the basis of miles driven also allowed for age, which is a key demographic in this study, to be used as a separate between-subjects variable. Previous research has used a variety of methods to define driving experience, with no clear and agreed upon method. A common method of defining driving experience is by how many years a person has possessed a driving license (e.g. Underwood, Crundall and Chapman, 2002; Crundall et al., 2011). Using this methodology, ‘inexperienced’ drivers are usually those who have possessed a driving licence for a few years at most, whereas ‘experienced’ drivers have held a licence for many years (the exact method varies from study to study). These methodologies may in part be due to the range of participants available in universities for research studies. However, a drawback to these methods is that they assume that all of the participants have been driving at a similar rate. There is no guarantee that drivers who are the same age or have possessed a driving license for the same amount of time have actually driven the same amount and therefore have the same amount of driving experience. Konstantopoulos, Chapman and Crundall (2010) utilised a similar method as used in this study, with novice drivers being categorised on the basis of the number of hours that they had driven.

In order to ensure that the participants who were not sure of their total mileage provided accurate driving mileages, the experimenter provided assistance by helping them calculate the average number of miles they drove each week and then extrapolating from that the total number of miles they had driven since getting their driving licence.

The participants were also divided into two groups depending on their age. Third year undergraduates were placed in the 'old' groups while first years were placed into the 'young' groups. This was viewed as the best way to get testing groups with the maximum difference in average age while utilising the relatively limited sample of participants available for testing within the university. This design led to the creation of four separate testing groups: 'old experienced', 'old novice', 'young experienced' and 'young novice'. Equal numbers of participants were placed in each these testing groups, with 8 participants in total assigned to each testing group. Males and females were equally represented in each of the four testing groups with four males and four females assigned to each testing group. This produced a fully counterbalanced design with regards to the age, experience and gender of the participants.

The average age of participants on the 'old' drivers was 21.41 years, with a range of 21 to 22 years. The average age of the 'young' drivers was 19.19 years, with a range of 18 to 20 years. The average driving experience of the 'experienced' drivers was 23,125 miles, with a range of 10,000 to 100,000 miles. The average driving experience of the 'novice' drivers was 1631 miles, with a range of 0 to 5000 miles. The average age of the male participants was 20.3 years with a range of 18 to 22 years. The average age of the female participants was 20.2 years, with a range of 18 to 22 years. The average driving experience of the male participants was 15,093 miles, with a range of 1 to 100,000 miles. The average driving experience of the female participants was 9,662 miles, with a range of 0 to 55,000 miles. Details of the demographics of the four testing groups can be found in Table 7.1.

Table 7.1. Demographic details for participant groups.

Group	Driving Experience (miles)		Age	
	Mean	Range	Mean	Range
Old Exp.	33,875	10,000 -100,000	21.5	21-21.7
Young Exp.	12,375	10,000-20,000	19.3	18.9-19.9
Old Novice	1,700	0-1,000	21.4	21-21.9
Young Novice	1,550	0-3,500	19.1	18.3-19.7

7.4.2 Materials

7.4.2.1 fNIR Device

The fNIR device that was used in Studies 3 and 4 was also used for this study. For further information on the fNIR device, please refer back to chapter 4 .

7.4.2.2. Driving Simulator

The experiment took place in the Nottingham Integrated Transport and Environment Simulation Facility (NITES) 2 simulator. The NITES 2 driving simulator consists of a stripped-down driving rig and a 180 degrees of vision projection screen. The rig consists of a seat, a steering wheel, a rear view mirror, instrument dial (speedometer and rev counter) and a gear lever. A large TV screen is placed behind the driving rig in order to present the participant with a view of the road behind them. Three projectors are used to project a simulated driving environment on to the 180 degrees projection screen.

7.4.2.3 Driving Simulations

5 driving scenarios, one for each driving condition, were created using custom driving simulation software (XPI Simulation, U.K.). The driving environment consisted of a 3-mile long single-carriageway road with a 60mph speed limit and hedges, trees, fields and the some buildings appearing on each side of the road. This is the exact same road that was used in Study 4. The 5 driving tasks in this study consisted of a following task in which the participants were instructed to follow the car in front and not attempt to overtake, as well as four overtaking tasks. The four overtaking tasks were differentiated by the density of traffic in both lanes of the single-carriageway road. The traffic densities range from the lowest density, Density 1, up to the highest traffic density, Density 4. In traffic Density 1, the opposite lane of the road was relatively empty providing ample opportunity to overtake. The traffic in the opposite lane steadily increased up until Density 4, whereby there was a steady stream of traffic in the opposite lane. As a result, the chances to overtake would be severely limited. In traffic density 4 (the highest density), approximately 83 cars passed by during the scenario if the participant drove at 60mph. In relation to this, traffic density 3 had 83%, traffic density 2 had 66% and traffic density 1 had 47% of the oncoming traffic present in study 4. The traffic on the participant's side of the carriageway was kept the same in all of the driving tasks in order to allow an equal number of opportunities to overtake in each task. The vehicles travelling in the same lane as the participant's simulated vehicles were programmed so that they drove at a speed of 50mph. This was done in order to help facilitate overtakes that did not require the participants to break the speed limit. All of the driving scenarios took place in dry, daytime road conditions.

From the participant's perspective, they are sitting in the driver's seat of a simulated car (which is right hand drive). They can see certain parts of the simulated car (in order to provide a reference point for placing the car on the road). This includes part of the front of the car and the rear window area of the back of the car (which is presented through the TV screen that provides a view of the road behind).

7.4.2.4 NASA TLX

The NASA TLX (NASA, Washington D.C., USA) workload scale was used to assess the subjective workload associated with each of the driving tasks. 6 scales are used to measure various aspects of workload. The scales are Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. Each scale is scored out of 21, with higher scores being an indicator of higher workload. The NASA TLX scale was modified by the addition of an inhibitory control measure (asking how much inhibitory control the participants believed they were using during the task) in order to establish the different levels of inhibition required in the different driving conditions. The inhibitory control question asked 'How much inhibitory control did you require to complete this task?'. Participants were required to given a score ranging from 1 (little inhibitory control required during the task) to 100 (a very large amount of inhibitory control required during the task). A 1-100 scale was chosen to match the format of the workload measures used in the NASA-TLX and to provide a quick measure of feedback that could be completed along side the NASA-TLX directly after each driving task. This method allowed for the inhibitory control associated with each driving tasks to be established, as well as the individual levels of inhibitory control for each participant. No piloting or validation was used in the creation of the inhibitory control scale.

7.4.2.5 Simulator Sickness Questionnaire

A standardised simulator sickness questionnaire was used to assess simulator sickness levels in the participants. It can be found in Appendix 3.1.

7.4.3 Design

A mixed design was implemented for this study, with both between-subjects and within-subjects variables. The dependant variables were: change in blood oxygenation, change in Hb, change in HbO_2, NASA TLX scale scores and the number of overtakes completed. The between-subjects independent variables were age ('old' and 'young'), driving experience ('experienced' and 'novice') and gender. The within-subjects independent variables were traffic density (with four levels) and driving task (following or overtaking).

The order of the driving conditions was counterbalanced so that all possible task orders were used in the experiment. The participants were randomly assigned to these task orders.

7.4.4 Procedure

The participants were first allowed to familiarise themselves with the driving simulator and its controls by completing a practise drive. This consisted of a drive along an empty three-mile long single carriageway road. Once the practise drive had been completed and the participant was comfortable with the simulator controls, the NIRS sensor was secured on their forehead (making sure that there was no hair in the way) using Velcro straps and a cloth (to help keep out external light). An fNIR reading was then taken to make sure that a clear signal was being received. The signal was checked by examining the raw fNIR signals displayed in the fNIR control software. The software displays the raw signals for the light at 730nm and 850nm. These signals must be within a range of 1000 to 4000mV in order to be regarded as valid and reliable signals. If this was not the case, the sensor pad was re-placed on the forehead as a bad signal often results from the sensor pad not being firmly secured to the forehead or hair obstructing the sensors. If the signal was still not at an appropriate level, the gain settings of the fNIR device were adjusted until the signal was within the acceptable boundaries.

Before beginning the driving tasks, fNIR baseline recordings were taken. This involved taking 10 seconds of fNIR recordings while the participants stared at the blank projection screen. Once the fNIR baselines were taken the experiment began. The fNIR and physiological devices were set to record at the exact moment that the first driving simulation/replay started. The participant was then instructed on what they were required to do in the upcoming task. For each of overtaking tasks the participants were instructed to overtake as many cars as they could while driving in a safe manner and obeying traffic laws. They were also instructed to attempt to drive at the speed limit of 60mph in order to ensure that that driving tasks lasted roughly the same length of time. For the following task the participants were instructed to follow the vehicle in front, while keeping a safe and constant distance, for the duration of the simulation. The driving simulation was then activated, while simultaneously the fNIR device began to record frontal lobe activity relative to the pre-task baseline. After the driving task had been completed, the simulation was shut down (leaving a blank projection screen) and the simultaneously fNIR device stopped recording frontal lobe activity. The participant then filled out the modified NASA TLX scale. After each hazard scenario had been completed a short break was taken during which time the participants completed the simulator sickness questionnaire (Appendix 3.1). If at any point scores of 'Moderate' sickness were reported then the experiment was paused and the participant was provided with some water. The participants were given several minutes after which they were asked if they wanted to continue. If they said that they felt unable to continue then they were removed from testing. If the participants stated that they now felt O.K. then the experiment continued. If at any point a participant reported a score of 'Severe' on any of the measures then they were provided with some water and they were removed from testing. A fan blowing cool air was left on at all times to help reduce incidences of simulation sickness. No participants were removed from testing due to simulator sickness in this study. This same procedure was followed for all five of the driving tasks.

Each driving task lasted approximately 3 minutes if the participants drove at the speed limit, meaning that each participant drove in the simulator for approximately 15 minutes in total if they drove at this speed. During the experiment, average speeds ranging from 50mph to 60mph were observed. The order of the tasks was fully counterbalanced between the participants. When combining the breaks taken in between each driving task to complete the NASA-TLX and the simulation sickness questionnaire, the total length of the experiment was approximately 25 minutes. If the participants crashed at any point during the driving tasks then the scenario was restarted unless the crash occurred very close to the end of the simulation.

7.5 Results

7.5.1 Number of Overtakes

A 2x2x2x4 (gender x age x driving experience x traffic density) mixed-subject ANOVA was used to analyse the effect of increasing traffic density on overtaking. The results revealed a main effect of traffic density $F(3,72)= 35.694$, $MSE= 2.958$, $p <.001$, ${}_{p}\eta^2 =.598$, showing that there were significant differences in the number of overtakes completed between the four traffic densities. A linear contrast revealed a significant linear trend $F(1,24)= 71.083$, $MSE=4.333$, $p <.001$, ${}_{p}\eta^2 =.748$, with increasing traffic density resulting in fewer overtakes. The analysis also revealed a main effect of gender $F(1,24)= 6.475$, $MSE=27.875$, $p <.05$, ${}_{p}\eta^2 =.212$, with males completing more overtakes than females. A significant interaction between gender and traffic density was also found $F(3,72)=3.190$, $MSE=2.958$, $p<.05$, ${}_{p}\eta^2 =.117$, which can be seen in Figure 7.1. A simple main effect analysis with a Bonferroni adjusted alpha level of .0125 for multiple comparisons revealed that the difference between males and females was significant for traffic Density 1 $F(1,96)=13.500$, $MSE=9.188$, $p<.001$, ${}_{p}\eta^2 =.123$ and traffic Density 3 $F(1,96)=4.408$, $MSE=9.188$, $p=.012$, ${}_{p}\eta^2 =.044$ only.

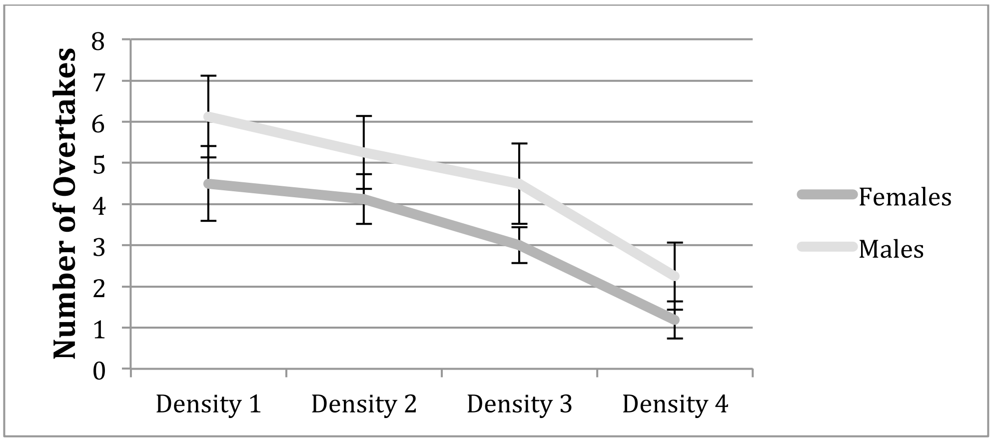


Figure 7.1. Significant interaction between traffic density and gender on the number of overtakes completed. Error bars represent +/- 1 S.E.

A summary of the overtaking performance from the different traffic densities, as well as the performance of different participant groups is summarised in Table 7.1.

Table 7.2. Mean number of overtakes completed in the four traffic densities by the different participant groups. Values in parentheses display the standard deviations.

Density	D1	D2	D3	D4
All Participants	6.5 (3.76)	4.9 (3.01)	4.1 (3)	2.1 (2.59)
Old	6.8 (3.64)	4.5 (2.42)	3.7 (1.73)	1.9 (1.81)
Young	6.1 (3.96)	5.2 (3.55)	4.5 (3.91)	2.2 (3.25)
Experienced	6.25 (4.1)	5.3 (3.75)	4.4 (3.83)	2.1 (3.29)
Novice	6.7 (3.53)	4.4 (2.06)	3.8 (1.94)	2.1 (1.75)
Male	8.4 (3.65)	5.6 (2.89)	5.2 (3.44)	3 (3.03)
Female	14.5 (2.75)	4.12 (3.03)	3 (2.03)	1.2 (1.72)

7.5.2 NASA TLX

The NASA-TLX data was first analysed using a 2x2x2x5x7 (gender x age x driving experience x driving task x scale) mixed ANOVA. Mauchly's Test of Sphericity was breached for scale and the scale by task interaction. As a result Greenhouse Geisser corrections were used. The results revealed a significant main effect of task $F(4,96)=63.426$, $MSE=4.872$, $p<.001$, ${}_{p}\eta^2=.725$, a significant main effect of scale $F(3.741,89.75)=36.316$, $MSE=45.259$, $p<.001$, ${}_{p}\eta^2=.602$ and a significant interaction between task and scale $F(7.764,186.324)=31.227$, $MSE=18.806$, $p<.001$, ${}_{p}\eta^2=.565$. The significant interaction between task and scale demonstrates that the scales differ in terms of the aspects of task workload and inhibition they are measuring. As a result of this interaction, individual ANOVAs for each of the NASA-TLX scales were carried out.

7.5.2.1 Following v Overtaking

The NASA TLX data was then analysed with using seven (one for each scale: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration, Inhibition) 2x2x2x2 (gender x age x driving experience x driving task) mixed-subjects ANOVAs in order to compare workload and inhibitory control scores in the following and overtaking driving conditions. For this analysis, data from the four traffic densities was averaged for each scale and compared to the scores from the following task.

A significant main effect of task for Mental Demand was found $F(1,24)=153.102$, $MSE=4.872$, $p<.001$, ${}_{p}\eta^2=.864$, with participants reporting that overtaking was more mentally demanding than following. There was also a significant main effect of gender for Mental Demand $F(1,24)= 6.291$, $MSE=6.977$, $p<.05$, ${}_{p}\eta^2=.208$, ${}_{p}\eta^2=.208$, with females finding both following and overtaking more mentally demanding than males. A significant interaction between driving task and experience $F(1,24)= 7.543$, $MSE=4.872$, $p<.05$, ${}_{p}\eta^2=.239$, was found and can be seen in can be seen in Figure 7.2. A simple main effects analysis with a Bonferroni adjusted alpha level of .025 for multiple comparisons revealed a significant main effect of experience in the following condition $F(1,48)=5.401$, $MSE=5.924$, $p<.05$, ${}_{p}\eta^2=.101$ but no significant effect in the overtaking condition, with novices finding the following task significantly less mentally demanding than experienced drivers.

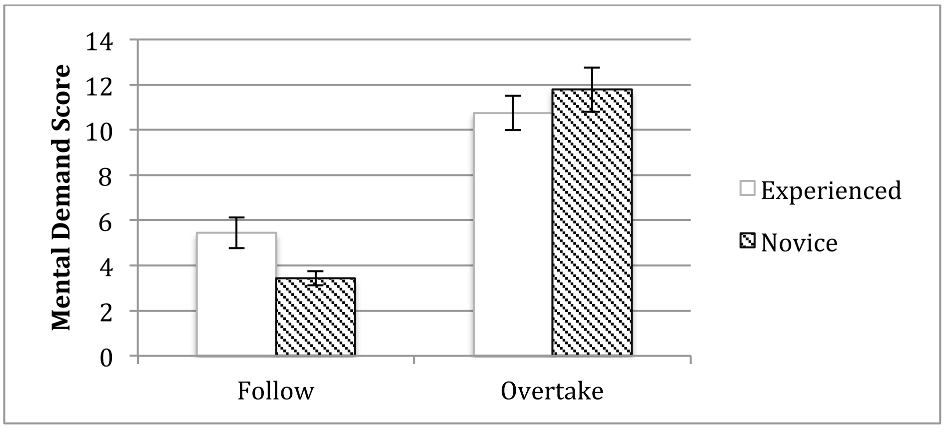


Figure 7.2. Significant interaction between driving task and driving experience for Mental Demand score. Error bars represent +/- 1 S.E.

A significant main effect of task for Temporal Demand was found $F(1,24)= 234.150$, $MSE=4.137$, $p<.001$, ${}_{p}\eta^2=.907$, with overtaking being more temporally demanding than following. There was also a significant interaction between task and experience for Temporal Demand $F(1,24)= 10.409$, $MSE=4.137$, $p<.05$, ${}_{p}\eta^2=.303$, which can be seen in Figure 7.3. A simple main effects analysis with a Bonferroni adjusted alpha level of .025 for multiple comparisons revealed that there was a significant difference between experience in the overtaking condition

$F(1,48)=11.005$, $MSE=6.958$, $p=.023$, ${}_{p}\eta^2=.187$, but no significant difference in the following condition. While novices and experienced drivers found the following task to be equally temporally demanding, novices found the overtaking task significantly more temporally demanding than experienced drivers.

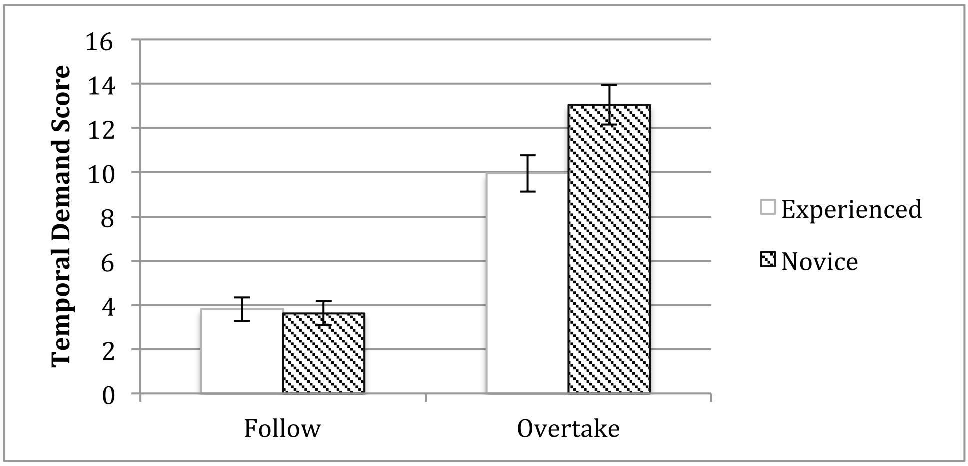


Figure 7.3. Significant interaction between driving task and driving experience for Temporal Demand score. Error bars represent +/- 1 S.E.

The analysis revealed a significant main effect of driving condition for Physical Demand $F(1,24)= 41.724$, $MSE=2.690$, $p<.001$, ${}_{p}\eta^2=.635$, with the participants reporting that overtaking was more physically demanding than following. There was also a significant main effect of driving condition for Performance $F(1,24)= 53.371$, $MSE=9.328$, $p<.001$, ${}_{p}\eta^2=.690$, with the participants reporting that they had performed better when following as opposed to overtaking. The Frustration scale also produced a main effect of driving task $F(1,24)= 63.403$, $MSE=9.516$, $p<.001$, ${}_{p}\eta^2=.725$ with overtaking being more frustrating than following. A significant main effect of driving condition for Effort was found $F(1,24)= 170.811$, $MSE=6.148$, $p<.001$, ${}_{p}\eta^2=.877$, with overtaking being rated as requiring more effort than following. A significant main effect of gender for Effort was also found $F(1,24)= 4.889$, $MSE=6.257$, $p<.05$, ${}_{p}\eta^2=.169$, with females noting higher levels of effort for both driving tasks compared to males.

A significant main effect of task for Inhibitory Control was found $F(1,24)= 179.445$, $MSE=6.986$, $p<.001$, ${}_{p}\eta^2=.882$, demonstrating that overtaking was perceived to require more inhibitory control than following. A significant main effect of age was also found for Inhibitory Control $F(1,24)= 5.000$, $MSE=10.970$, p

<.05, ${}_{p}\eta^2$=.172, with the older drivers reporting higher levels of inhibitory control in both driving tasks compared to the younger drivers. A further significant main effect of driving experience was found for Inhibitory Control $F(1,24)= 5.519$, $MSE=10.970$, $p <.05$, ${}_{p}\eta^2$=.187, with experienced drivers reporting higher levels of inhibitory control than novice drivers. A significant interaction between age and experience for Inhibitory Control was also found $F(1,24)= 6.830$, $MSE=10.970$, $p <.05$, ${}_{p}\eta^2$=.222, and can be seen in Figure 7.4. A simple main effects analysis with a Bonferroni adjusted alpha level of .025 for multiple comparisons revealed that there was a significant main effect of age for the experienced drivers $F(1,24)=11.759$, $MSE=10.970$, $p=.019$, ${}_{p}\eta^2$ =.331 with experienced older drivers reporting higher inhibitory control than younger experienced drivers in both driving conditions. There was no significant main effect of age for novice drivers however.

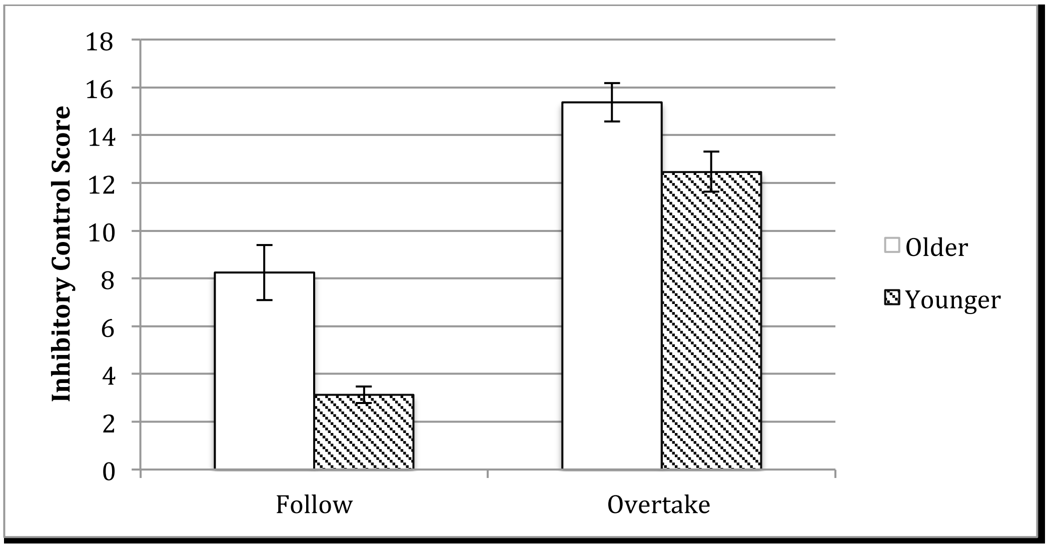


Figure 7.4. Significant interaction between age and experience on inhibitory control scores. Error bars represent +/- 1 S.E.

A summary of the differences between NASA TLX scores between the following and overtaking conditions can be found in Table 7.3.

Table 7.3. Average NASA TLX scores for the following and overtaking conditions. Overtaking means were calculated by averaging the scores from the four traffic densities. Values in parentheses display the standard deviations.

Scale	Following	Overtaking
M.D.	4.44 (2.31)	11.27 (3.53)
P.D.	2.91 (1.55)	5.55 (3.18)
T.D.	3.72 (2.10)	11.50 (3.76)
Perf.	11.86 (5.03)	17.44 (3.38)
Eff.	4.47 (2.49)	12.57 (3.33)
Frust.	3.72 (2.30)	9.86 (4.95)
I.C.	4.41 (3.54)	13.26 (3.71)

7.5.2.2 Effect of Traffic Density

To investigate the effect of traffic density on NASA TLX scores, seven (one for each scale: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration, Inhibition) 2x2x2x4 (gender x age x driving experience x traffic density) mixed-subjects ANOVAs were carried out.

The ANOVAs showed no significant effects for Mental Demand, Physical Demand, Effort and Frustration. There was a significant main effect of experience for Temporal Demand $F(1,24)= 8.460$, $MSE=36.203$, $p <.05$, ${}_{p}\eta^2 =.261$, with novices reporting higher temporal demand across the four traffic densities. There was also a significant main effect of traffic density on Performance $F(3,72)= 15.592$, $MSE=9.840$, $p<.001$, ${}_{p}\eta^2 =.394$. A linear contrast revealed a linear trend $F(1,24)= 46.906$, $MSE=9.642$, $p<0.001$, ${}_{p}\eta^2 =.662$, showing that as traffic density increased, perceived driving performance decreased. There was also a significant main effect of traffic density for Inhibitory Control $F(3,72)= 3.080$, $MSE= 6.110$, $p<.05$, ${}_{p}\eta^2 =.114$. A linear contrast revealed a linear direction $F(1,24)= 5.994$, $MSE=6.757$, $p<.05$, ${}_{p}\eta^2 =.200$, with participants recording higher levels of inhibitory control as the traffic density increased.

A summary of the NASA TLX scores from the different traffic densities can be seen Table 7.3.

Table 7.4. Average NASA TLX from the four traffic densities. Values in parentheses display the standard deviations.

Density	D1	D2	D3	D4
M.D	10.88 (3.06)	11.22 (3.67)	11.63 (3.32)	11.34 (4.12)
P.D.	5.25 (2.78)	5.41 (3.16)	5.72 (3.20)	5.84 (3.63)
T.D.	11.19 (3.18)	11.88 (3.70)	11.50 (4.26)	11.44 (3.95)
Perf.	9.09 (5.90)	11.25 (5.03)	12.97 (4.04)	14.13 (3.54)
Effort	11.66 (2.75)	12.56 (3.46)	13.03 (3.26)	13.03 (3.74)
Frust.	8.56 (4.81)	10.19 (4.55)	10.38 (5.39)	10.31 (5.03)
I.C.	12.19 (3.72)	13.59 (3.65)	13.28 (3.48)	13.97 (3.90)

7.5.3 fNIR Analysis

After all of the fNIR data had been collected, it became apparent that two of the fNIR channels were not working correctly. As a result, data from these two channels was excluded from the analysis (leaving data from 14 fNIR channels to analyse). For one participant, data from two of the five driving conditions was missing. As these missing values represented only 1.25% of all of the data, they were replaced with weighted cell means. In order to remove any erroneous data from 'bad' channels, data that fell beyond 2 standard deviations (positive or negative) was removed and replaced with the participant's global condition means. For oxygenation, 2.5% of values were replaced. For Hb, 2.7% of values were replaced. For HbO_2, 3.1% of values were replaced.

Pre-processing of the fNIR data involved averaging the Hb and HbO_2 values for each of the 14 channels over time (the entire duration of the driving task). Hb and HbO_2 values were used to calculate the corresponding changes in blood oxygenation (HbO_2 – Hb) throughout the driving tasks. These values were also averaged for each of the 14 channels over the duration of the task. All of the blood values are changes relative to baseline and are measured in micro mols (μM).

The fNIR data was averaged over time in order to get an overall, global view of the changes in haemodynamic activity throughout the driving tasks. During the driving tasks there were multiple points of interest and differences in behaviours and processes required. The following task differed from the overtaking task in that no decision-making was required. There were also differences in motor control between the following and overtaking conditions as the following task simply required the participants to follow the car in front, whereas the overtaking tasks requiring precise and accurate control inputs for any overtakes. These differences in decision-making and motor control were also present between the different overtaking densities. As the traffic density increased, the decision-making process on whether to overtake or not became more difficult. The increasing traffic density also meant that overtakes needed to be completed more quickly and accurately to avoid any collisions. It was decided that encapsulating all of these factors into the analysis was important. Providing individual analyses for all of these events would not have been practical.

The decision to average over time was also made as the experimental setup made it very difficult and time-consuming to provide accurate synchronisation between events in the simulator and the fNIR data. There were also difficulties in finding comparable events that could be studied across all or most of the participants. For example, participants differed significantly in the number of overtakes they made or attempted, as well as the time they took to do these things. To balance this out, a further analysis that focused on individual points of interest was also made.

Mauchly's Test of Sphericity was breached for channel for every fNIR analysis. Therefore Greenhouse-Geisser corrections were used to calculate the degrees of freedom.

The fNIR data was first analysed using a 2x2x2x5x14 mixed ANOVA for the change in blood oxygenation, Hb and HbO_2. Mauchly's Test of Sphericity was breached for Channel so Greenhouse Geisser corrections were used.

There was a significant difference in blood oxygenation changes between the five driving tasks $F(4, 96)= 3.427$, $MSE=13.001$, $p <.05$, ${}_{p}\eta^2 =.125$. A planned Helmert contrast revealed that the following task was significantly different from the four overtaking tasks $F(1, 24)= 13.569$, $MSE=11.580$, $p <.01$, ${}_{p}\eta^2 =.361$ but there were no other significant differences. There were no significant effects of age [$F(1,24)= 0.267$, $MSE=82.785$, $p =.610$, ${}_p\eta^2 =.011$], driving experience [$F(1,24)= 0.01$, $MSE=82.785$, $p =.922$, ${}_p\eta^2 =.000$] or gender [$F(1,24)= 1.385$, $MSE=82.785$, $p =.251$, ${}_p\eta^2 =.055$]. There was a significant main effect of channel $F(4.703, 112.8)= 3.852$, $MSE=2.662$, $p <.001$, ${}_p\eta^2 =.186$ demonstrating that blood oxygenation changes were higher at some channels than others.

There was no significant difference in the change in HbO_2 between the five driving tasks [$F(4, 96)= 1.692$, $MSE=8.041$, $p =.158$, ${}_p\eta^2 =.066$]. There were also no significant effects of age [$F(1,24)= 0.033$, $MSE=53.303$, $p =.858$, ${}_p\eta^2 =.001$], driving experience [$F(1,24)= 0.000$, $MSE=53.303$, $p =.993$, ${}_p\eta^2 =.000$] or gender [$F(1,24)= 1.40$, $MSE=53.303$, $p =.248$, ${}_p\eta^2 =.055$]. There was a significant main effect of channel $F(4.348,104.354)= 0.110$, $MSE=6.335$, $p <.05$, ${}_p\eta^2 =.088$ demonstrating that the change in HbO_2 was higher at some channels than others.

There was a significant difference in the change in Hb between the five driving tasks $F(4, 96)= 3.484$, $MSE=14.257$, $p <.05$, ${}_p\eta^2 =.127$. A planned Helmert contrast revealed that the following task was significantly different from the four overtaking tasks $F(1, 24)= 10.448$, $MSE=2.901$, $p <.01$, ${}_p\eta^2 =.303$ but there were no other significant differences. There were no significant effects of age [$F(1,24)= 1.841$, $MSE=1.378$, $p =.187$, ${}_p\eta^2 =.071$], driving experience [$F(1,24)= 0.059$, $MSE=1.378$, $p =.810$, ${}_p\eta^2 =.002$] or gender [$F(1,24)= 0.319$, $MSE=1.378$, $p =.578$, ${}_p\eta^2 =.013$]. There was a significant main effect of channel $F(4.109,159.844)= 4.105$, $MSE=0.732$, $p <.001$, ${}_p\eta^2 =.146$ demonstrating that the change in Hb was higher at some channels than others.

7.5.3.1 Following vs. Overtaking

In order to directly compare fNIR data between the following task and the overtaking tasks, fNIR measures (blood oxygenation, Hb and HbO_2) from the four overtaking conditions were averaged across one another to produce 14 channels of data for each participant that represented all of the overtaking traffic densities. Analysis of the fNIR measures for the overtaking and following tasks took the form of 2x2x2x2x14 (gender x age x driving experience x driving task x channel) mixed-subject ANOVAs. No voxel group comparisons were possible as a result of the missing channels.

7.5.3.1.1 Blood Oxygenation

The results of the ANOVA revealed a significant main effect of task for blood oxygenation $F(1,24) = 11.666$, $MSE=5.023$, $p < .05$, ${}_p\eta^2 = .327$, with higher levels of blood oxygenation when overtaking compared to following. There was a significant main effect of channel $F(3.954,94.901) = 5.646$, $MSE=1.416$, $p < .001$, ${}_p\eta^2 = .190$, with significant differences in the change in blood oxygenation between the 14 channels. There were no significant effects of age, [$F(1,24) = 0.632$, $MSE=30.633$, $p = .434$, ${}_p\eta^2 = .026$], driving experience [$F(1,24) = 0.103$, $MSE=30.633$, $p = .751$, ${}_p\eta^2 = .004$], or gender [$F(1,24) = 1.211$, $MSE=30.633$, $p = .202$, ${}_p\eta^2 = .048$].

7.5.3.1.2 HbO_2

Although not significant [$F(1,24) = 3.885$, $MSE=19.663$, $p = 0.06$, ${}_p\eta^2 = .301$], HbO_2 levels were higher in the overtaking task than in the following task. There was no significant main effect of channel [$F(4.650,111.605) = 1.509$, $MSE=0.795$, $p=.112$, ${}_p\eta^2 = .059$]. There were no significant effects of age [$F(1,24) = 0.343$, $MSE=18.895$, $p = .564$, ${}_p\eta^2 = .014$], driving experience [$F(1,24) = 0.084$, $MSE=18.895$, $p = .775$, ${}_p\eta^2 = .003$] or gender [$F(1,24) = 0.815$, $MSE=18.895$, $p = .376$, ${}_p\eta^2 = .033$].

7.5.3.1.3 Hb

There was a significant main effect of task for Hb $F(1,24) = 10.349$, $MSE=15.186$, $p < .01$, ${}_p\eta^2 = .301$, with higher levels of Hb in the following task than the overtaking task. There was a significant main effect of channel $F(3.030,72.716) = 3.646$, $MSE=1.076$, $p < .05$, ${}_p\eta^2 = .132$, with significant differences in the change in blood oxygenation between the 14 channels. There were no significant effects of age [$F(1,24) = 2.515$, $MSE=2.469$, $p = .126$, ${}_p\eta^2 = .095$], driving experience [$F(1,24) = 0.201$, $MSE=2.469$, $p = .658$, ${}_p\eta^2 = .008$] or gender [$F(1,24) = 2.694$, $MSE=2.469$, $p = .114$, ${}_p\eta^2 = .101$].

7.5.3.2 Effect of Traffic Density

To compare DLPFC activity between the four different traffic densities, 2x2x2x4x14 (gender x age x driving experience x traffic density x channel) mixed-subject ANOVAs were used for blood oxygenation, Hb and HbO_2. Mauchly's Test of Sphericity was breached for traffic density for Hb. Therefore Greenhouse-Geisser corrections were used to calculate the degrees of freedom.

7.5.3.2.1 Blood Oxygenation

There was no significant difference between blood oxygenation across the four traffic densities [$F(3, 72)= 1.230$, $MSE=14.257$, $p =.305$, ${}_p\eta^2 =.049$]. There were no significant effects of age [$F(1,24)= 0.110$, $MSE=74.168$, $p =.743$, ${}_p\eta^2 =.005$], driving experience [$F(1,24)= 0.112$, $MSE=74.168$, $p =.730$, ${}_p\eta^2 =.005$] or gender [$F(1,24)= 1.409$, $MSE=74.168$, $p =.247$, ${}_p\eta^2 =.055$]. There was a significant main effect of channel $F(4.109,98.606)= 3.852$, $MSE=3.654$, $p <.01$, ${}_p\eta^2 =.138$ demonstrating that blood oxygenation levels were higher at some channels than others.

7.5.3.2.2 HbO_2

There was no significant difference in the levels of HbO_2 between the four traffic densities [$F(3,72)= 0.969$, $MSE=8.016$, $p =.412$, ${}_p\eta^2 =.039$]. There were no significant effects of age [$F(1,24)= 0.003$, $MSE=51.524$, $p =.955$, ${}_p\eta^2 =.000$], driving experience [$F(1,24)= 0.024$, $MSE=51.524$, $p =.878$, ${}_p\eta^2 =.001$] or gender [$F(1,24)= 1.676$, $MSE=51.524$, $p =.208$, ${}_p\eta^2 =.065$]. There was a significant main effect of channel $F(4.082,97.957)= 2.522$, $MSE=3.322$, $p <.05$, ${}_p\eta^2 =.095$, demonstrating that HbO_2 levels were higher at some channels than others.

7.5.3.2.3 Hb

There was no significant difference in the levels of Hb between the four traffic densities [$F(1.640,39.358)= 2.004$, $MSE=3.740$, $p =.115$, ${}_p\eta^2 =.078$]. There were no significant effects of age [$F(1,24)= 1.223$, $MSE=7.128$, $p =.280$, ${}_p\eta^2$

=.048], driving experience [$F(1,24)= 0.363$, $MSE=7.128$, $p =.552$, $_p\eta^2 =.015$] or gender [$F(1,24)= 0.004$, $MSE=7.128$, $p =.950$, $_p\eta^2 =.000$]. There was a significant main effect of channel $F(5.390,129.353)= 2.634$, $MSE=0.977$, $p <.05$, $_p\eta^2 =.099$, demonstrating that Hb levels were higher at some channels than others.

Summaries of the fNIR data can bee seen in Tables 7.4 and Table 7.5.

Table 7.5. Mean blood oxygenation, Hb and HbO_2 values for following and overtaking analysis. Values in parentheses display the standard deviation.

	Following	Overtaking
Blood Oxy.	-0.225 (1.157)	0.367 (1.087)
Hb	0.099 (0.42)	-0.16 (0.356)
HbO_2	-0.09 (0.891)	0.203 (0.893)

Table 7.6. Mean change in blood oxygenation*, Hb and $Hb0_2$ values and standard deviations for the four traffic densities.*

Density	D1	D2	D3	D4
Oxygenation				
mean	0.599	0.126	0.387	0.358
sd	1.434	1.753	1.246	1.113
Hb				
mean	-0.21	-0.005	-0.311	-0.115
sd	0.469	0.95	0.391	0.436
HbO2				
mean	0.396	0.15	0.1	0.167
sd	1.281	1.01	1.17	0.893

7.5.4 Overtaking Event Analysis

In order to investigate whether increased haemodynamic activity during the active overtaking task was primarily due to inhibition or overtaking, an analysis was carried out in which periods during which the participants were overtaking other vehicles was compared with periods in which overtaking was not possible due to the presence of vehicles in the opposite lane. The logic of this analysis was that periods in which the participants were overtaking would be periods during which task workload would be at its highest, whereas the periods in which no overtaking was possible would be periods in which inhibition would be at its highest. Overtaking is likely to result in the highest levels of workload as the participants are required to make the decision to overtake on the basis of the road environment and then use motor skills to complete the overtake without crashing. No overtaking periods are likely to result in the highest levels of inhibition as the participants must inhibit the impulse to attempt an overtake due to it not being safe to do so. The mean change in blood oxygenation during the overtaking and no overtaking periods was selected as the haemodynamic measure as it provides a combined measure of the changes in Hb and HbO_2 throughout these events. Data from all of the participants who successfully completed at least one overtake was included in this analysis. Due to differences in how the participants drove, there were differences in the number of overtakes completed, the time spent overtaking and the time spent where overtakes were not possible, so as much data as possible was included from each participant in the analysis. Overtaking periods were defined as the point at which the participants fully committed to an overtaking manoeuvre (i.e. the point at which the participants had fully crossed into the next lane to attempt an overtake), until they had re-joined their lane and carried on driving ahead. This process usually took approximately 10 seconds. No overtaking periods were defined as periods of time in which an overtaking manoeuvre was impossible due to there being a vehicle in the opposite lane that was too close to allow for an overtake to be made. Data was taken from periods of at least 10 seconds in which overtaking was not possible in order to provide a more comparable measure with the overtaking periods.

The change in blood oxygenation was averaged over all 14 fNIR channels. Blood oxygenation was averaged over each of the 14 fNIR channels in order to provide a straightforward and general view of how overall haemodynamic activity changed during the overtaking manoeuvres and no overtake possible phases. This decision was also made to help reduce any noise in the data that may have been present due to taking data from a relatively short period of time across all 14 of the fNIR channels. The decision to average across the fNIR channels was also taken due to two of the fNIR channels having been removed from the analysis due to them having been faulty.

As all of the participants completed at least one overtake in the lowest traffic density, it was possible to analyse data from this density using a mixed ANOVA that included the between subjects measures. A 2x2x2x2 (driving period x age x experience x gender) mixed ANOVA revealed a significant main effect of driving phase $F(1,24)= 4.918$, $MSE=1.244$, $p <.05$, ${}_{p}\eta^2 =.170$, with significantly higher levels of blood oxygenation when the participants were overtaking compared to when they were unable to overtake. There were no other significant main effects or interactions. The data from this analysis can be seen in Table 7.6.

Table 7.7. Mean change in blood oxygenation (measured in μ mols) for the overtaking and no overtaking periods for all participants in traffic density 1.

Driving Period	No Overtaking	Overtaking
Change in Oxygenation		
mean	-0.071	0.205
sd	0.177	0.246

For traffic densities 2, 3 and 4 it was not possible to carry out ANOVAs including the between subject variables as not all of the participants completed at least one overtake and therefore there were not equal numbers from each of the age, experience and gender groups. As a result of this, t-tests were used to compare the overtaking and no overtaking periods. A t-test comparing the overtaking and no overtaking periods for traffic density 2 revealed no significant difference in the mean change in blood oxygenation [t(22)=0.994, p=.331] between the overtaking (M=0.07, S.D.=1.46) and no overtaking periods (M=-0.047, S.D.=1.32). A t-test comparing the overtaking and no overtaking periods for traffic density 3 revealed no significant difference in the mean change in blood oxygenation [t(22)=0.374, p=.712] between the overtaking (M=0.135, S.D.=1.25) and no overtaking periods (M=0.099, S.D.=1.15). A t-test comparing the overtaking and no overtaking periods for traffic density 4 revealed no significant difference in the mean change in blood oxygenation [t(22)=1.587, p=.127] between the overtaking (M=0.672, S.D.=1.81) and no overtaking periods (M=0.362, S.D.=1.29).

Table 7.7 lists the mean change in blood oxygenation during the overtaking and no overtaking periods for each of the participants included in the analysis across the four traffic densities.

Table 7.8. Mean change in blood oxygenation (measured in μ mols) for the overtaking and no overtaking periods for the 23 participants who completed at least one overtake in each of the four traffic densities.

	Driving Period	No Overtaking	Overtaking
Change in Oxygenation			
Density 1	mean	-0.104	0.221
	sd	0.205	0.312
Density 2	mean	-0.047	0.072
	sd	0.275	0.306
Density 3	mean	0.099	0.135
	sd	0.241	0.262
Density 4	mean	0.362	0.673
	sd	0.269	0.376

In order to compare the overtaking and no overtaking periods between the four different traffic densities, data from 23 participants who were able to complete at least one overtake in each of the traffic densities was compared using a 2x4 (traffic density x driving period) ANOVA. Mauchly's test of Sphericity was breached for driving period and the driving period by density interaction. Greenhouse Geisser corrections were therefore used to calculate the degrees of freedom. A 2x4 (driving period x traffic density) ANOVA revealed no significant main of driving period [F(1,22)= 0.937, MSE=3.669, p =.344, ${}_{p}\eta^2$ =.041], demonstrating that there was no significant difference in the mean change in blood oxygenation between the overtaking and no overtaking periods. There was also no significant main effect of traffic density [F(3,66)= 1.061, MSE=1.083, p =.372, ${}_{p}\eta^2$ =.046], demonstrating that the mean change in blood oxygenation was not significantly different between the different traffic densities. There was also no significant interaction between driving period and traffic density [F(1.551,34.113)= 1.698, MSE=1.922, p =.202, ${}_{p}\eta^2$ =.072].

7.6 Discussion

The results revealed that blood oxygenation in the DLPFC increased during the overtaking tasks compared to the following task. HbO_2 levels showed a non-significant increase while Hb levels were lower in the overtaking tasks than in the following task. These results are in line with the subjective workload and inhibition scores in which the participants reported higher levels of workload and inhibition when overtaking. However no effects of age, driving experience or gender were uncovered for frontal lobe activity when comparing overtaking with following. This is despite the fact that these factors affected the subjective workload and inhibitory control scores when comparing the following and overtaking tasks. The study also failed to find differences in any of the measures of frontal lobe activity due to changes in traffic density, as well as any effect of age, experience or gender across the different densities. This is despite the fact that subjective workload and inhibition scores, as well as the number of overtakes showed some effect from changes in traffic density.

The increased levels of blood oxygenation and the decreased levels of Hb in the overtaking task indicate that DLPFC activity was higher when the participants were overtaking compared to following. The increase in blood oxygenation for overtaking compared to following is in line with the results from study 4, which showed that blood oxygenation was higher in the active overtaking task than in the active following task. The decreases in Hb in the overtaking tasks are similar to what happened in Study 4, where Hb levels were lower in the active overtaking task than the active following task. This increase in DLPFC activity was mirrored by the subjective workload and inhibition scores, which were higher for overtaking compared to following. Overtaking resulted in higher inhibition scores as well as being more mentally demanding, physically demanding, temporally demanding, frustrating, effortful as well as being inferior in terms of performance. The subjective workload scores are very similar to those found in Study 4. The subjective workload and inhibition scores indicate that the increased DLPFC activity for overtaking was due to the increased workload and inhibition that overtaking produced compared to following.

The failure of the fNIR data to show differences due to age and driving experience may be related to the sample of participants recruited in this study. The range of ages and driving experience in this study was potentially too limited. Research from the likes of Sowell et al. (1999b) has made it clear that the frontal lobes do not reach full maturation until about 25 years of age (the age at which fatality rates in drivers begin to fall significantly (ECMT, 2006)). In this study the 'young' driver group had an average age of 19 years. It is possible that, in order to observe differences in frontal lobe activity, younger drivers who are 17 years of age (the age at which driving becomes legal) may need to be tested. Research has shown than drivers at this age are more likely to be involved in crashes than drivers who are slightly older (Forsyth, 1992). It also appears that the effects of age and experience on driving are not linear and have the greater effect the younger drivers are, with this effect having even been seen between 16 and 17 year old drivers (Maryland et al. 2001). This may be due to the changes in the physiology of the frontal lobes, and the accompanying cognitive and emotional changes are most prominent in the very youngest drivers. Jongen et al. (2011) found that that inhibitory control was lower in the 17-18 year olds than 22-24 year olds, showing that there are differences in the executive functions in people of this age. This reduced inhibitory control was accompanied by increased risk taking. This may explain why drivers in their late teens demonstrate an even greater propensity to be involved in accidents than young adults.

In the 'old' group in this study the average age was about 21. It may be necessary to test drivers that are older than this, namely those who are closer to 25 as this is when the frontal lobes are fully maturated. As mentioned above, the effects of age and driving experience are rapid and significant in young drivers, with those aged 17-20 involved in far more accidents than those aged 20-24 (Forsyth, 1992), with a rapid falloff in fatality rates for those aged 25 to 64 (ECMT, 2006). Executive functions are also undergoing significant changes at this age (Jongen et al. 2011). By testing a group of drivers between the ages of 17 and 25 any differences in frontal lobe activity may become apparent. It may also necessary to compare the frontal lobe activity of 17-25 year olds with older and more experienced adult drivers who have long had fully matured frontal lobes and have significant amounts of driving experience. Given the interactions between age and experience on driving, recruiting participant groups that do not have a wide enough range of ages may mask some of the effects experience has. This will provide a detailed look into the DLPFC activity of drivers who are undergoing maturational changes in the frontal lobes as well as those who have fully maturated frontal lobes.

A limitation of the study may be related to individual differences in the participants. In particular, individual differences in impulsivity may have impacted the results regarding the between-subjects variables. For example, it is possible that some of the participants in this study may have been naturally more impulsive than other participants, which could have affected how the participants responded to and performed in the different driving tasks. As mentioned previously, research has shown that impulsivity can lead to increased risk taking (Steinberg, 2007; Steinberg, 2008). Impulsivity may also impact on the way that people drive with impulsivity being associated with dangerous driving behaviours (Williams, 1998; Lee, 2007). Unfortunately there was no measure in place in this study to account for any possible differences in the impulsivity of the participants. In order to improve the study, measures of impulsivity could be taken and then factored in to the between-subjects comparisons. This could be achieved by using a measure such as the Barratt Impulsiveness Scale (Barratt, 1965; Patton, Stanford and Barratt, 1995). This method would make it possible to determine if differences or a lack of differences between the groups was actually due to differences in impulsivity rather than differences in age, experience or gender.

Although no differences were found in frontal lobe activity as a function of age or experience, the subjective workload and inhibition scores do indicate that there were differences between these groups. Novices were found to find following less mentally demanding while finding overtaking more temporally demanding. Both the older and the more experienced participants reported higher levels of inhibitory control for the overtaking tasks, while the older experienced participants reported higher levels of inhibitory control for both following and overtaking. As executive functions have been found to develop with age (Weinberger et al. 2005; Jongen et al. 2011), the finding that the older participants reported higher levels of inhibitory control might be a reflection of them having more developed executive functions than the younger participants. The steep decline in accident rates in young drivers may in part be due to increased executive functions such as inhibition. The higher levels of inhibitory control from the more experienced participants could be due experience teaching drivers to be more aware of dangers on the road, and as a result they are more likely to inhibit potentially dangerous driving behaviours. The fact that the older experienced drivers reported higher levels of inhibitory control for both following and overtaking could indicate an interaction between driving age and experience that effects executive control while driving. This is in line with research that has shown crash rates reduce as a function of both age and experience (Forsyth et al., 1995). It is also possible that the results indicate that the main driver of DLPFC activity in these driving tasks is not inhibition but workload. More evidence for this hypothesis can be taken from the analysis of the overtaking and no overtaking periods.

Inspection of the mean change in blood oxygenation during the overtaking and no overtaking periods provides a mixed picture that tends to lean towards the idea that task workload, rather than inhibition, is the primary factor behind DLPFC activity in these driving tasks. The results from traffic density 1 are similar to those from the active overtaking task in study 4 (which used the same task and traffic density), with the overtaking periods leading to significantly higher changes in blood oxygenation than the no overtaking periods. This finding suggests that workload is the primary driver of increased blood oxygenation in the DLPFC during the overtaking tasks. However, this result is not replicated in traffic densities 2, 3 and 4 where no significant differences were found between the overtaking and no overtaking periods. A possible explanation for this finding is that as the traffic density increased so did the level of task inhibition, whereas the level of task workload remained relatively unchanged. As a result of this the changes in blood oxygenation during the no overtaking periods may have increased relative to the overtaking periods. This could mean that task inhibition was also contributing to DLPFC oxygenation. It is also important to note that as the traffic density increased the number of overtakes decreased, meaning that there was less data to sample from.

The finding that there was no significant difference in the change on blood oxygenation between the four traffic densities (for either the overtaking or no overtaking periods) fits with the finding that the four traffic densities did not differ in terms of the changes in blood oxygenation across the entire length of the tasks. Given that the purpose of increasing the amount of traffic on the road was to increase the level of task inhibition required whilst keeping the level of task workload roughly the same, this result would again suggest that task workload is the most important factor for DLPFC activity.

The failure to find significant differences between the traffic densities (both in overall terms and between the overtaking and no overtaking periods) and the failure to find differences between the overtaking and no overtaking periods for densities 2, 3 and 4 would appear to provide contradictory results. The failure to find differences between the traffic densities would indicate that inhibition is not contributing to DLPFC activity whereas the failure to find differences between the overtaking and no overtaking phases may indicate that inhibition is contributing to DLPFC activity. This is further complicated by the fact that there were significant differences between the overtaking and no overtaking periods in traffic density 1. One possible explanation for this is that inhibition is having an effect but that this effect is limited. This limited effect may be enough to result in there being no significant difference between the overtaking and no overtaking periods (combined with the reduced number of overtakes to sample from) but may not be great enough to produce significant differences between the different overtaking densities. Overall, these results indicate that workload is primarily driving DLPFC activity, however task inhibition may also contribute to DLPFC activity.

As with age and experience, gender was not found to have any effect on frontal lobe activity. This is despite the fact that females reported finding both the following and overtaking tasks more mentally demanding and effortful. There was also a significant difference in the number of overtakes completed, with males overtaking more than females. Although the driving behaviours, offence rates and accident rates of males and females have been found to be different (Harre et al. 1996), it is not entirely clear why this is. Research has suggested that it is due to increased risk taking and sensation seeking in males. As controlling these behaviours has been linked to frontal executive functions (Weinberger et al. 2005), it was predicted that differences between male and female drivers would also manifest themselves in terms of frontal lobe activity. However, research has not specifically investigated differences in frontal lobe activity between male and female drivers. The difference between male and female drivers is likely due to a wide range of complex issues such as a peer pressure, societal expectations and personality factors (Geraghty and Shah, 2010). Therefore it is perhaps not surprising that the differences between male and female drivers did not manifest in frontal lobe activity differences.

By decreasing the average number of overtakes completed, increasing the traffic density had the desired effect on driving behaviour by making overtaking more difficult. However, increasing traffic density was not found to have any significant effects on frontal lobe activity. This is despite the fact that the participants reported that their levels of inhibition increased through the traffic densities. Subjective performance levels also decreased indicating that the participants found the task more difficult as the traffic density increased. No effects for age or experience were found for the different traffic densities either. However, in this case the only difference in the subjective workload and inhibition scales was that novices reported higher temporal demand across the driving densities. As increasing traffic densities was designed to primarily increase levels of inhibition in the overtaking tasks, these results could again provide evidence that the predominant factor in frontal lobe activity in the overtaking tasks is workload and that the difference in workload between the traffic densities was not large enough to produce significant differences in frontal lobe activity. This is supported by the fact that several of the workload scales, including Mental and Physical Demand, did not show significant increases across the traffic densities. Another factor to consider is task difficulty. It is possible that as the traffic density increased beyond a certain point the participants found it too difficult to overtake. The number of overtakes completed and perceived performance both showed reductions as a result of increasing traffic density, supporting the idea that the participants found the tasks more difficult. This may have led to the participants changing the way that they approached the tasks, namely giving up in attempting to overtake. This would impact on the levels of workload and inhibition in the task. No significant effects for gender were found across the traffic densities, which was in line with the subjective workload and inhibition scores that showed no gender differences.

The NASA-TLX data revealed that subjective measures of task inhibition did increase across the traffic densities while only two measures of task workload did. The failure to witness these patterns as well as a general increase in blood oxygenation throughout the overtaking manoeuvre could again indicate that workload is primarily responsible for DLPFC activity in these tasks or the issues result from differences in task difficulty.

It is important to note that there were limitations in the method selected to assess the subjective levels of inhibitory control in the driving tasks. The measure used was not a pre-established measure of inhibitory control and did not undergo any validation or pilot testing. An inhibitory control questions styled on the format of the NASA-TLX was chosen for primarily for practical reasons. The measure used allowed for the data to be collected in the same 0-100 scale format as is used in the NASA-TLX. This measure also allowed for data to be collected quickly during the time that the participants were completing the NASA-TLX and for the participants to report their subjective levels of inhibitory control for the individual driving tasks directly after they had finished those tasks. It is possible that the inhibitory control measure was not providing a valid and reliable measure of inhibitory control. Using an established measure of inhibitory control, adapted to address levels of inhibitory control within individual driving tasks could help overcome this issue.

Studying the number of overtaking manoeuvres completed provides an insight into driving behaviours of the different groups. The results revealed that the male drivers completed more overtake than the females. The literature on gender differences in driving has indicated that males are far more likely to be involved in crashes than females (Forsyth, 1992; Wells et al. 2008) and display more risk taking behaviours when driving (Harre et al. 1996). The act of overtaking another car can be a very risky process, with the possibility of colliding with another car or crashing when returning to the correct lane. It is therefore possible that the increased number of overtakes by males is an indication that they are willing to take more risks and are maybe less aware of the risks from overtaking than the female participants. A further explanation for why females carried out fewer overtakes than males may be related to the fact that the female participants reported finding the driving tasks more mentally demanding and effortful than males. If the female participants found the tasks more demanding and difficult the differences in overtaking may simply have been due to this fact. The overtaking data shows that although there were no differences in frontal lobe activity between males and females, there were differences in driving behaviour along with the subjective workload scores.

To summarise, the aim of this study was to investigate whether or not age, experience and gender had any effect on DLPFC activity, subjective workload and inhibition, as well as driving behaviour (the number of overtakes completed). A further aim was to verify the results from Study 4 and further investigate the role that workload and inhibition play in driving through increasing traffic density across four overtaking tasks. The subjective workload and inhibition scores, as well as the number of overtaking manoeuvres showed changes across the driving conditions and between the different participant groups. DLPFC activity was found to be higher during overtaking than following. This supports the findings from Study 4. However no significant changes in frontal lobe activity were found across the four overtaking traffic densities. This is despite the fact that inhibition showed increases across the four traffic densities (which is what the increased traffic densities were primarily designed to do), whereas many of the workload measures failed to show differences. This may have been indicative of workload being the primary driving force behind frontal lobe activity in these tasks or have been due to task difficulty.

No effects of age or experience were found when comparing overtaking with following or when comparing the different overtaking traffic densities. The failure to find these differences may have been due to the sample of participants not having a wide enough range of ages and experiences. The failure to find differences in frontal lobe activity for gender was perhaps not unexpected, as previous literature has not detailed differences in DLPFC activity between males and females. Differences in subjective workload and driving behaviour between males and females were observed however. It is also possible that overtaking and following tasks are not the best driving tasks to investigate the role of the frontal lobes in driving and differences between groups of drivers. Research has shown that age, experience and gender affect driving in a number of ways including in terms of hazard perception abilities (Lee, 2007).

In order to better understand how DLPFC is operating in these tasks and how workload and inhibition contribute to the changes in haemodynamic activity, several changes could be made to the methodology in order to improve this study. One of the limitations to this study related the difficulties in synchronising the fNIR data to events of interest such as overtakes. There were also difficulties in identifying comparable non-overtaking moments and having overtakes that could be compared between the participants (as some participants failed to complete any overtakes while others differed in the number of overtakes completed). A change that could be made to this study would be to alter the driving scenarios to help synchronise events of interest in the driving tasks to the fNIR recordings. The behaviour of the AI vehicles in the overtaking tasks could be altered so that there was a guarantee for at least one overtaking chance to occur in each of the density conditions. The timings if these events could also be controlled, which would also allow for easier synchronisation with the fNIR recordings. This would make it much easier to tie events of interest, such as an overtaking manoeuvre, to any corresponding haemodynamic changes. The same process could be used to guarantee that at certain other times overtaking manoeuvres would not be possible.

An alternate solution could be to change the nature of the driving tasks themselves to something that allows for easier synchronisation and identification of comparable points of interest. For example, a driving task in which participants are required to drive across a set of roundabouts would involve similar decision-making and motor control skills as an overtaking task. The participants would have to determine when it is safe to pull out and when it is not, much like in the overtaking tasks where they had to decide whether or not it was safe to overtake. The points at which the participants successfully pull out and cross the roundabouts could be compared to the points at which they are forced to wait to cross the roundabout. This could provide a way to compare moments where workload is the predominant factor (crossing the roundabout) with times in which inhibition may be the main factor (waiting to cross the roundabout). The amount of traffic using the roundabout could be increased to mirror the effect of the increase in traffic density for the overtaking tasks. However this time, the tasks could be designed in such a way as to guarantee that at least one successful roundabout crossing could be completed in each traffic density, which was not the case in the overtaking studies.

Chapter 8 - The DLPFC and Hazard Perception

8.1 Abstract

In order to establish if the DLPFC plays a role in the process of hazard perception (HP), a set of simulated driving HP driving tasks were used in conjunction with fNIR. Participants either completed the simulated HP task that was used in study 2, or they completed a HP task in which they watched replays of those eight simulated hazards in the simulator. Both tasks required the participants to respond to any driving hazards using a button response. fNIR was used to monitor DLPFC activity during the HP tasks. The results revealed that the participants made significantly more button responses and received significantly higher scores for the replay hazards than the simulated hazards. However, contrary to expectations, no significant differences in DLPFC activity were found between the two hazard tasks. A possible explanation for this finding is that the majority of the task workload was associated with the process of detecting the hazards, rather than the physical aspects of the task (i.e. the process of driving through the simulated environments or simply watching replays of the hazards). A further explanation may be that the similar visual environment provided by the replay hazards provided a more realistic driving experience and in the process increased task workload in line with the simulated hazards. Overall the results indicate that the use of replays may provide a means with which to better imitate real driving for the purposes of HP testing.

8.2 Introduction

Building upon the findings of this thesis regarding HP and the role of the frontal lobes in driving, this chapter will bring those research strands together by monitoring DLPFC activity during simulated HP testing. The purpose of study 2 was to compare responses between video and simulated driving hazards in order to determine if the video-based method used in most HP tests is representative of real world HP and if this method can be improved in any way to better mimic HP while driving on the road. This chapter will continue this research by comparing different HP testing formats as well as comparing frontal lobe activity between these different testing formats.

The button press method and the simulated HP driving scenarios from Study 2 will be used again but this time they will be compared to pre-recorded replays of those driving simulation, rather than the video hazards used in study 2. The purpose of this is to provide a more comparable visual experience across both experimental conditions, removing the possibility that a difference in the visual method of presenting (i.e. video on a T.V. or projected simulation) the hazards may impact on the results. The use of replay hazards as opposed to video hazards may also provide a method of introducing simulated elements into standard video-based HP tests in order to provide an improved test of HP ability. Comparing simulated hazards with replays of those hazards also removes any possible effect that differences in the content of the HP videos and simulations may have on responses, which was a potential issue in study 2.

A further change is that this time, instead of recording physiological measures such as skin conductance, fNIR will be used to monitor DLPFC activity during the simulated and replay hazards. This will be done in order to assess whether or not there are differences in DLPFC activity between the replay and simulated hazards. This will provide an insight into the relative levels of task workload for replay and simulated hazards. As the replay hazards are similar in nature to the video hazards used in the UK HP test, while the simulated hazards are being used as a means of more closely mimicking on-road driving, the fNIR findings have the ability provide more detail on whether or not HP testing methods in which people simply watch and respond to hazards can be representative of HP while driving on the roads.

8.3 Study 6

The aim of Study 6 is to further explore simulated HP testing by monitoring DLPFC activity with fNIR as participants complete one of two simulated HP tests. Two different HP tests will be used: a simulated driving HP test (the same as was used in Study 2), and a replay HP test consisting of watching and responding to replays of the simulated drives (which will be projected in the driving simulator). As was the case in study 2, responses will be made through the use of steering wheel mounted button. The number of button presses per second and the scores for each individual hazard will be recorded. The predictions for Study 6 are that simulated driving will result in significantly higher levels of DLPFC activity, as well as fewer button responses and lower HP scores than the replay hazards. These predictions are based upon the findings of Study 2, in which simulated hazards were associated with higher physiological responses, fewer button responses and lower hazard scores than the video hazards. One interpretation of those results was that the extra workload induced by actually controlling the vehicle was impairing hazard perception performance. In the current study we can explore this possibility by directly measuring workload through DLPFC activity in each task.

8.4 Methodology

8.4.1 Participants

A power analysis revealed that for a large effect size (F = 0.4), 48 participants (24 in each group) would be required for this study.

48 participants (29 females, 19 males) took part in the study. All were students at the University of Nottingham and all of the participants had full driving licences. All of the participants had normal or corrected-to-normal vision. The average age of the participants was 24.6 years, with a range of 18 to 28 years. Average driving experience was 5.25 years with a range of 6 months to 10 years. The mean number of miles driven per year was 6032.27, with a range of 100 to 10,000 miles. All of the participants had normal or corrected-to-normal vision.

8.4.2 Apparatus and Stimuli

8.4.2.1 Driving Simulator

The experiment took place in NITES2 (the Nottingham Integrated Transport and Environment Simulation facility) driving simulator. This is a fixed base driving simulator with a 180 degrees semi-circular projection screen 2 metres in height and 5 metres in diameter. A rear view was provided by a flat screen television placed behind the participant that could be viewed with a conventional rear view mirror (no side mirrors were present). This simulator provides all standard car controls, with the exception of a handbrake. Hazard responses were made via a pushbutton attached to the steering wheel (for both replay and driven hazards).

8.4.2.2 Driving Simulations (Driven Hazards)

The exact same 8 HP driving simulations (based upon 8 training clips provided by the DVSA) that were used in Study 2 were used in this study. Hazard scores for the simulated hazards were also calculated using the same method as was used in Study 2 (i.e. via fixed coordinates in the simulated road environment and a Matlab script that extracted the time that the participants crossed these points in space), with participants having to press a button on the steering wheel when they perceived something hazardous in the simulated environment.

8.4.2.3 Driving Replays (Replay Hazards)

8 HP driving replays were used in this study. They consisted of recordings of drives through each of the 8 simulated driving scenarios that were then projected onto the 180 degrees projection screen. The participants sat in the driving simulator, but instead of driving through the simulated environment, they simply watched the replays and responded to hazards using the response button located on the steering wheel. These recorded drives were designed to be as representative of a person driving through the simulations as possible. This involved sticking to the speed limit and the other rules of the road, driving in as 'normal' and safe a way as possible. By doing this the scoring hazards in the replays triggered as they were designed to, with them coming into view at the right time and the simulated car reacting in a safe manner to avoid the hazard. This allowed for accurate scoring windows, defined by set time-points in the replays that signalled the onset and offset of the scoring windows (e.g. the scoring window for hazard 1 starts 47.6s into replay and ends 51s into replay). The total length of the replays and the length of the hazard windows are similar to those of the respective video hazards used in study 2. The length of the hazard replays ranged from x to x (mean), while the length of the scoring windows ranged from x to x (mean). Detailed timings for the replay hazards can be found in Table 8.1.

Table 8.1. List of total hazard clip length and scoring window lengths for each of the eight replay hazards.

Hazard	Total Clip Length (s)	Hazard Window Length (s)
1	60.01	4.5
2	75.89	10.8
3	60.98	3.3
4	52.74	3.6
5	64.02	4.1
6	75.07	9.5
7	64.98	6.31
8	65.84	13.14

8.4.2.4 fNIR Device

The fNIR device that was used in Studies 3, 4 and 5 was also used for this study. For further information on the fNIR device, please refer back to chapters 4 and 5.

8.4.2.5 Simulator Sickness Questionnaire

A standardised simulator sickness questionnaire was used to assess simulator sickness levels in the participants. It can be found in Appendix 3.1.

8.4.3 Design

A mixed design ANOVA was utilised for this study, with the between-subjects factor being the driving condition (simulated vs. replay). The within subjects factor was the 8 different simulated or replay hazards. The first dependent variable of interest was reaction time after hazard onset for participants to press the response button. This was converted into a score, with early responses receiving 5, and later responses getting lower values (corresponding to the standard DVSA testing procedure). The total number of button presses inside and outside of the hazard scoring windows was recorded to provide further dependent variable related to perceptions of danger – this was the number of button responses (per second) made during each HP video or simulation. The final dependent variable related to the fNIR recordings, with the change in oxygenation, Hb and HbO_2 across the 16 fNIR channels during the simulated or replay tasks being recorded.

8.4.4 Procedure

All of the participants began by completing a practice drive in the driving simulator that allowed them to familiarise themselves with the driving controls of the simulator as well as understanding the HP task they required to complete in the experiment. Although the participants in the replay condition did not need to drive in the simulator they still completed a practice drive. The purpose of this was to make sure that both groups of participants received the same training. After the practice drive the participants were free to ask any questions they had about the experiment. The participants in the replay condition were instructed to sit in the simulator, watch the driving scenarios that were projected in the simulator and press the steering wheel mounted button when the perceived there to be a driving hazard present. After each hazard scenario had been completed a short break was taken during which time the participants completed the simulator sickness questionnaire (Appendix 3.1). If at any point scores of 'Moderate' sickness were reported then the experiment was paused and the participant was provided with some water. The participants were given several minutes after which they were asked if they wanted to continue. If they said that they felt unable to continue then they were removed from testing. If the participants stated that they now felt O.K. then the experiment continued. If at any point a participant reported a score of 'Severe' on any of the measures then they were provided with some water and they were removed from testing. A fan blowing cool air was left on at all times to help reduce incidences of simulation sickness. No participants were removed from testing due to simulator sickness in this study. It took approximately 15 minutes to complete the replay hazard tasks, including the breaks taken in-between each of the replay hazards to complete the simulator sickness questionnaire.

For the simulated hazards the participants were instructed to follow the road ahead unless instructed to turn left or right (which was communicated through a pre-recorded message that was set to play at specific locations during the simulated drives). The participants were also instructed to take avoiding action to avoid having a collision with any objects in the road as well as to respond to any hazards using the steering wheel mounted response button. The participants were also informed of the speed limit for each simulated hazard and were instructed to try and drive at the speed limit. After each hazard scenario had been completed a short break was taken during which time the participants completed the simulator sickness questionnaire. If the participants drove at the speed limit then the simulated hazard condition lasted for approximately 20 minutes, including the breaks taken between each of the eight simulated hazards to complete the simulator sickness questionnaire. If the participants crashed at any point during one of the simulated hazards then the scenario was restarted unless the crash occurred very close to the end of the simulation.

Once the participant fully understood the task, the fNIR sensor was secured on their forehead (making sure that there was no hair in the way) using Velcro straps and a cloth (to help keep out external light). An fNIR reading was then taken to make sure that a clear signal was being received. The signal was checked by examining the raw fNIR signals displayed in the fNIR control software. The software displays the raw signals for the light at 730nm and 850nm. These signals must be within a range of 1000 to 4000mV in order to be regarded as valid and reliable signals. If this was not the case, the sensor pad was re-placed on the forehead as a bad signal often results from the sensor pad not being firmly secured to the forehead or hair obstructing the sensors. If the signal was still not at an appropriate level, the gain settings of the fNIR device were adjusted until the signal was within the acceptable boundaries.

Before beginning the driving tasks, fNIR baseline recordings were taken. This involved taking 10 seconds of fNIR recordings while the participants stared at the projection screen (which was blank at the time). Once the fNIR baselines were taken the experiment began. The fNIR was set to record at the exact moment that the first driving simulation or replay started and ended at the conclusion of the final simulation or replay. For the 8 driven HP scenarios, the procedure was exactly the same as the procedure used in Study 2. The participant's task was to follow a set route in the scenarios and react to any hazards in the environment through the driving controls (e.g. braking or steering in order to avoid having a collision) as well as by responding by pressing the steering wheel mounted button.

For the replay condition, the participants simply had to sit in the driving simulator and watch the 8 replays, responding to hazards using the steering wheel mounted button. Both the simulated drives and the replays were completed in the order listed in Table 8.1.

8.5 Results

8.5.1 Hazard Scores

Hazard scores were analysed using a 2x8 (condition x hazard) mixed ANOVA. The results revealed a significant main effect of hazard $F(7,322)= 9.708$, $MSE=27.908$, $p <.001$, ${}_{p}\eta^2 =.174$, with some hazards receiving significantly higher scores than other. There was also a significant main effect of condition $F(1,46)= 17.891$, $MSE=67.503$, $p <.001$, ${}_{p}\eta^2 =.280$, with significantly higher hazard scores in the replay condition compared to the simulated condition. Table 8.2 provide a detailed breakdown of the hazard scores.

Table 8.2. Mean hazard scores for the 8 replay and simulated hazards. Values in parentheses are standard deviations.

Condition	Replay	Simulated
Hazard		
1	2.5 (1.4)	1.5 (1.6)
2	3.1 (1.8)	1.7 (1.7)
3	4.1 (1.1)	3.1 (1.8)
4	3.5 (1.2)	2.7 (1.9)
5	2.7 (1.8)	2.2 (2)
6	4.2 (0.8)	3.7 (1.6)
7	3.8 (1.2)	3.2 (1.8)
8	3.1 (1.4)	2.2 (1.8)

8.5.2 Button Presses (Per Second)

Button presses per second were analysed using a 2x2x8 (condition x window x hazard) mixed ANOVA. Mauchly's Test of Sphericity was breached for hazard and the window by hazard interaction. As a result Greenhouse Geisser corrections were used for the degrees of freedom. A significant main effect of condition was found $F(1,46) = 13.304$, $MSE = 0.734$, $p < .05$, ${}_{p}\eta^2 = .224$, with significantly more button presses per second for the replay hazards compared to the simulated hazards. A significant main effect of window was found $F(4.482,125.5) = 392.481$, $MSE = 13.168$, $p < .001$, ${}_{p}\eta^2 = .895$, with significantly more button presses per second taking place during the hazard window than outside of the window. There was also a significant main effect of hazard $F(4.047,186.166) = 12.966$, $MSE = 0.315$, $p < .001$, ${}_{p}\eta^2 = .220$, with some hazards being responded to significantly more than others. Table 8.3 provides a detailed breakdown of button presses. A significant interaction between window and hazard was also found $F(7,194.958) = 10.781$, $MSE = 0.258$, $p < .001$, ${}_{p}\eta^2 = .190$, which can be seen in Figure 8.1.

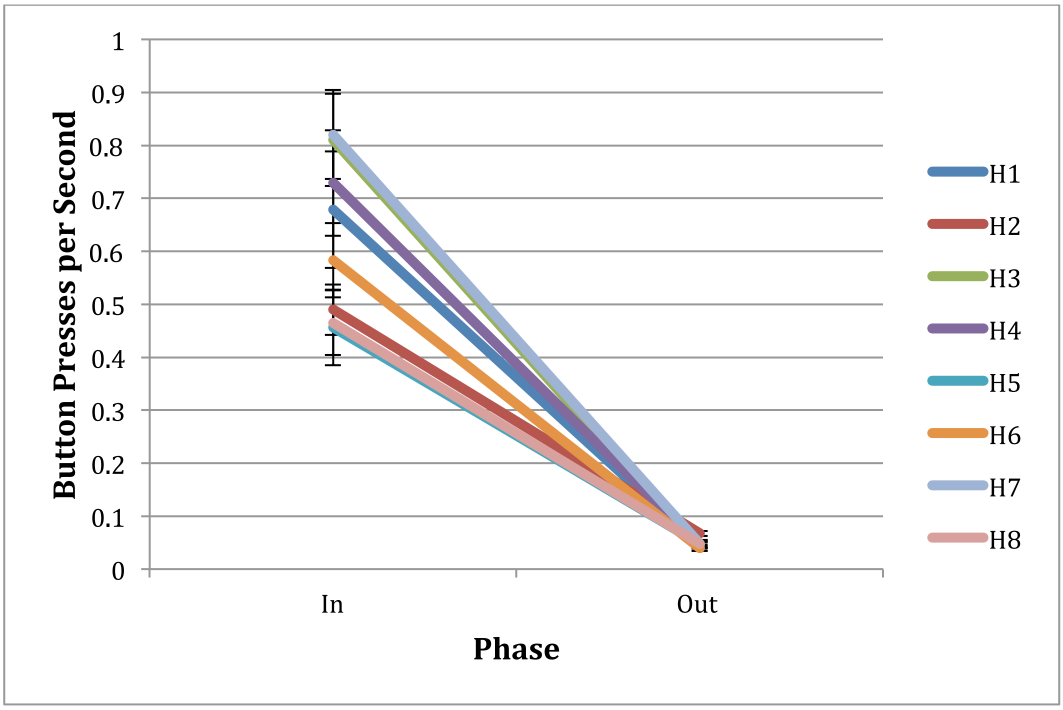


Figure 8.1. The hazard by phase interaction for button presses per second. Error bars represent +/- 1 S.E.

Table 8.3. Mean button presses per second for the 8 replay and simulated hazards. Values in parentheses are standard deviations.

Condition	Replay		Simulated	
Window	In	Out	In	Out
Hazard				
1	0.38 (0.21)	0.05 (0.02)	0.29 (0.32)	0.04 (0.03)
2	0.2 (0.12)	0.03 (0.02)	0.12 (0.1)	0.01 (0.02)
3	0.5 (0.19)	0.04 (0.03)	0.3 (0.2)	0.03 (0.03)
4	0.4 (0.2)	0.05 (0.02)	0.31 (0.3)	0.04 (0.03)
5	0.25 (0.15)	0.05 (0.03)	0.2 (0.19)	0.04 (0.04)
6	0.33 (0.17)	0.04 (0.02)	0.25 (0.17)	0.03 (0.03)
7	0.49 (0.19)	0.05 (0.01)	0.3 (0.21)	0.04 (0.04)
8	0.29 (0.14)	0.05 (0.03)	0.17 (0.15)	0.04 (0.04)

8.5.3 fNIR

The fNIR data was averaged over time in order to get an overall, global view of the changes in haemodynamic activity throughout the driving tasks. During the driving tasks there were multiple points of interest and differences in behaviours and processes required. It would not have been possible to create analyses for all of these points of interest for practical and technical reasons. The change in oxygenation, Hb and HbO_2 were analysed using 2x8x16 (condition x hazard x channel) ANOVAs. Mauchly's Test of Sphericity was breached for hazard across each of the measure. As a result, Greenhouse Geisser corrections were used for the degrees of freedom.

8.5.3.1 Oxygenation

A significant main effect of hazard was found $F(2.205,101.439)= 8.501$, $MSE=122.514$, $p <.001$, ${}_{p}\eta^2 =.156$, with the change in blood oxygenation greater for some of the hazards than others. Details of changes in blood oxygenation can be found in Table 8.4. No significant main effect of condition was found

[F(1,46)= 0.032, MSE=3.410, p =.858, $_p\eta^2$ =.001]. No other significant main effects or interactions were found.

Table 8.4. Changes in blood oxygenation for the 8 replay and simulated hazards.

Condition	Replay	Simulated
Hazard		
1	-0.13 (0.9)	-0.24 (1.3)
2	-0.04 (1.3)	0.14 (1.4)
3	0.008 (1.6)	0.19 (1.5)
4	0.01 (1.3)	0.1 (1.2)
5	0.33 (1.3)	0.4 (1.5)
6	0.38 (1.4)	0.39 (1.5)
7	0.34 (1.5)	0.43 (1.5)
8	0.4 (1.5)	0.43 (1.6)

8.5.3.2 HbO_2

A significant main effect of hazard was found F(2.065,190.555)= 18.646, MSE=254.609, p <.001, $_p\eta^2$ =.288, with the change in blood oxygenation greater for some of the hazards than others. Details of the changes in HbO_2 can be found in Table 8.4. No significant main effect of condition was found [F(1,46)= 0.653, MSE=67.555, p =.423, $_p\eta^2$ =.014]. No other significant main effects or interactions were found.

Table 8.5 Changes in HbO_2 for the 8 replay and simulated hazards.

Condition	Replay	Simulated
Hazard		
1	-0.03 (0.9)	-0.4 (1.3)
2	0.08 (1)	-0.01 (1.5)
3	0.2 (1.2)	-0.03 (1.5)
4	0.25 (0.8)	0.05 (1.2)
5	0.38 (1.2)	0.03 (1.4)
6	0.49 (1.4)	0.46 (1.4)
7	0.62 (1.5)	0.46 (1.7)
8	0.65 (1.6)	0.47 (1.7)

8.5.3.3 Hb

A significant main effect of hazard was found $F(2.885,132.729)= 5.762$, $MSE=17.708$, $p <.05$, ${}_{p}\eta^2 =.111$, with the change in blood oxygenation greater for some of the hazards than others. Details of the changes in HbO_2 can be found in Table 8.4. No significant main effect of condition was found [$F(1,46)= 1.254$, $MSE=89.758$, $p =.269$, ${}_{p}\eta^2 =.027$]. No other significant main effects or interactions were found.

Table 8.6 Changes in Hb for the 8 replay and simulated hazards.

Condition	Replay	Simulated
Hazard		
1	0.1 (0.6)	-0.21 (0.9)
2	0.12 (0.7)	-0.2 (1.2)
3	0.17 (0.8)	-0.16 (1.3)
4	0.25 (0.8)	0.05 (1.2)
5	0.2 (0.9)	0.14 (1.3)
6	0.25 (0.9)	0.04 (1.5)
7	0.24 (0.9)	0.04 (1.5)
8	0.28 (0.9)	0.001 (1.6)

8.6 Discussion

The results of study 6 revealed that, as expected, replay hazards received significantly more button responses and significantly higher hazard scores that the simulated hazards. There were also significant differences in the number of button presses and hazard scores that the eight different hazards received. However, contrary to expectations, no significant difference in frontal lobe activity was found between the replay and simulated hazards. Significant differences in frontal lobe activity were found between the different hazards however, matching the differences in the behavioural data for the hazards.

The findings that replay hazards were associated with significantly more button presses and higher hazard score was in line with prediction and with the findings of study 2. One of the potential explanations is that the process of detecting hazards is more difficult when driving, compared to when simply watching hazards. If detecting the simulated hazards were more difficult then we would expect the participants to have responded less and received lower scores for those hazards. As simulated hazards are being used as an analogue of real world driving, the implication of this reason for the response difference is that simply watching and responding to hazards may not provide a representative test of the HP skills needed when driving on the roads.

As with study 2, an alternate explanation for these differences relates to the dual task nature of the simulated hazard task. The participants completing the simulated hazards were tasked with driving through the simulated environments and avoiding colliding with any hazards, as well as responding to any hazards with a button press. The participants in the replay condition were only given one task, to respond to any hazards with a button press. It is possible that the lower hazard scores and fewer button responses to the simulated hazards stemmed from the participants being preoccupied with operating the driving simulator controls as opposed to any actual differences in HP ability and difficulty. The primary task in the replay task was to respond to any hazards with a button press, however the primary task in the simulated driving task was to use the simulator controls to drive through the simulated environment. It is therefore plausible that the participants in the simulated task were prioritising driving through the simulated environment as opposed to responding to any hazards with a button press. For example, when approaching a hazard the participants in the simulated condition may have been focused on avoiding having a collision with a hazard, resulting in them responding fewer times and later to that hazard compared to the participants in the replay condition.

Contrary to expectations DLPFC activity was not significantly higher for the simulated hazards. This prediction was also based upon the findings of study 3, 4 and 5, which indicated that the major factor behind increased DLPFC activity recorded by fNIR was the task workload. The fact that there was no significant difference in DLPFC activity between the simulated and replay conditions therefore indicates that the task workload was roughly equal across both conditions. Given that the replay and simulated tasks differed in their physical requirements (i.e. the simulated hazards required use of the simulator controls as well as button presses) but both required HP with the same eight hazards, it would appear that the primary workload factor in both tasks was the HP requirement.

In turn this finding would suggest that the replay HP test was able to evoke similar levels of workload and cognitive processing as simulated driving. As the replay hazards were similar in nature to the video-based HP test used in the UK HP test (in terms of simply watching and responding to hazards with a button press), while the simulated hazards were designed to provide an analogue of on-road driving, this result may suggest that video-based HP tests provide a good analogue for HP during real world driving with respect to the mental workload and cognitive processes required for HP during driving. The findings may also suggest that replay hazards provide a means of helping to bridge the gap between video-based HP and on-road HP by introducing simulated elements into the UK HP test without the need for full-on implementation of driving simulations. The fNIR results also lend support to the possibility that the reason for the fewer button responses and lower hazard scores for the simulated hazards was due to the dual-task associated with the simulated hazards. If detecting and responding to the simulated hazards was more difficult then we might have expected that this would have led to increased task workload and increased DLPFC activity for the simulated hazards. Instead it seems more likely that the reduced rate of responding in the simulated condition is related to output interference – the fact that drivers may delay responding using the button press until they have successfully negotiated the hazard. While this implies that responses to simulated hazards might have to be scored differently to those from watched hazards, it does not suggest that the two scenarios involve fundamentally different processes

One area in which significant differences in DLPFC activity occurred was between the eight hazards, with significant differences in hazards scores and the number of button presses also found (which mirrors the findings of study 2 when comparing simulated with video hazards). These results indicate that the individual hazards differed in the levels of workload and the number of perceived hazards that they contained, as well as differing in how difficult the scoring hazards were to detect and respond to in time. This is not surprising given that the content of the hazard simulations and replays differed across the eight hazards. Some of the simulations and their replays contained more potential hazards (e.g. other vehicles on the road and pedestrians on the pavements). These potential hazards may have elicited button presses, increased frontal lobe activity as the participants processed the potential danger and hazardousness of these vehicles and pedestrians, as well as affecting how accurate the participants were in responding to the scoring hazards.

Hazard 7 produced some of the largest increases in frontal lobe activity and button presses, whereas hazard 2 was associated with some of the lowest levels of frontal lobe activity and button presses. In hazard 7 a motorbike pulls out into the road heading towards you from a side street on the left and then proceeds to pass a parked car on the side of the road by moving into your lane. In hazard 2 a car emerges from a motorway slip road and into the left hand lane of the motorway in front of your car. Hazard 7 contains a number of potential hazards that the participants may have been reacting to, including: a cyclist in the opposite lane, other cars pulling out from junctions and other cars turning into side roads. Hazard 2 on the other hand does not contain such potential hazards, with cars on the other side of the road separated by a central reservation and the cars on the driver's side of the road maintaining a relatively large distance out in front. This makes it likely that many of the participants were responding to the scoring hazard and not any of the other vehicles in the road.

When inspecting hazard scores, we can see that hazard 1 received the lowest scores while hazard 6 received the highest. In hazard 1 a white van pulls out into the road in front of the car from a left hand side street. In hazard 6, while following a cyclist who has come from a side street, a parked car pulls out and forces the cyclist to swerve to avoid it. It is possible that the presence of the cyclist in hazard 6 made the participants more alert to potential hazards, in turn aiding reaction times to the scoring hazard. In hazard 1 meanwhile, there was no such factor that might have increased the overall awareness of the participants. The scoring hazard also required participants to scan ahead quite far as the white van first appeared a fairly significant distance away. Overall, the differences between the hazards demonstrate that the methodology used was sensitive to differences in the driving scenarios and hazards, which does lend support to the button press methodology for use in HP tests, whether they be video or simulated.

Given the finding in study 2 that simulated hazards were associated with higher levels of skin conductance than the video hazards, we may have expected the simulated hazards in this study to have produced higher levels of DLPFC activity than the replay hazards (which, like the video hazards in study 2, require the participants to simply watch and respond to the hazards). One possible explanation for this difference is that the replay hazards provide a closer experience to simulated hazards than the video hazards. The major difference between the replay HP test and the video HP test is the visual environment, with the replay hazards being projected onto the 180 degrees projection screen whereas the video hazards were presented on a T.V. screen. The different visual stimulation from the high fidelity driving simulations may have had an influence on the workload associated with the tasks. Ultimately, in order to better understand this issue, a further study could be carried out in which video, replay and simulated hazards are directly compared with one another using fNIR. Skin conductance could also be recorded to help establish if it correlates with DLPFC activity recorded with fNIR.

Overall, the fNIR finding suggest that the major workload component to the simulated driving task was not the operation of the driving simulator controls but the detection of hazards, which was required for both the simulated and the replay hazards. HP being responsible for the majority of the workload in the tasks would help explain the similar levels of DLPFC activity in the simulated and replay conditions. The behavioural results meanwhile are likely a reflection of the methodology used for testing HP performance.

As well as exploring what role the frontal lobes may play in HP, study 6 also aimed to investigate different HP testing methods in order to see if there are any suitable alternatives to the video-based method used in the UK HP test. In order to provide a direct comparison between the testing methods used here and the UK HP test, a button press methodology was chosen. While this method allows for this comparison and provides an easy means through which to grade HP performance it does come with the drawback of creating a double task for the simulated hazards. Ultimately this issue makes the accuracy of the HP scoring (i.e. whether or not the scores are an accurate reflection of HP ability as opposed to a response issue) debatable.

One way of resolving this issue would be to grade HP performance on the basis on simulator control inputs. Steering inputs, braking timings, braking pedal pressure, acceleration and speeds could be used to grade HP performance. For example, applying pressure to the brake pedal or changing the angle of the steering wheel could be used as an indication that a hazard has been detected and the candidate is taking appropriate avoiding action. However, due to the greatly different system of measuring HP performance, implementing this method into the UK HP test would require each HP simulation (based on the HP videos) to be validated to determine how to assign the correct score to the simulator inputs. Unfortunately, this would be a time consuming and expensive process, potentially limiting the attractiveness of this method.

Chapter 9 – General Discussion

9.1 Overview

The main aim of this thesis was to explore what factors affect HP performance, both in standard video-based testing environments as well as during simulated driving. This was achieved through comparing and evaluating HP performance in filmed and simulated driving environments, as well as through the exploration of actual data from the candidates taking the test to better understand what factors affect performance on the official UK HP test. A further aim of this thesis was to explore the psychophysiology of driving. This involved recording physiological responses and frontal lobe activity (measured with fNIR) across a range of driving tasks (including HP) that were either video-based or simulated. The overall purpose of this research is to improve safety on the roads

Chapter 2 of this thesis investigated what factors can influence HP test performance through an examination of candidates' button presses during the actual test, while Chapter 3 compared performance on video and simulator-based HP tests. The focus of the thesis then switched to the use of fNIR to explore the role of the frontal lobes in driving and HP. Chapter 5 used a set of classical frontal cognitive tasks to validate the fNIR device used in this thesis, while Chapters 6 and 7 used fNIR to investigate what role the frontal lobes play in driving. Chapter 8 combined the fNIR and HP elements of this thesis together by using fNIR to investigate what role the frontal lobes play in HP. This discussion chapter will review the findings of the studies carried out in this thesis as well as providing an in-depth discussion on HP testing and fNIR based upon the findings of this thesis. The real-world applications of this research and potential for future research will also be discussed.

9.2 Summary of Findings

Study 1 had the goal of creating a HP typology in order to investigate if there are any underlying factors that influence performance on the official UK HP test (e.g. are certain types of hazards better at discriminating between those who pass and fail the HP test), with the aim of exploring if there were any improvements that could be made to the test in its current form. Data from the UK HP test was analysed with a multidimensional scaling in order to create the HP typology. The DVSA provided one month's worth of data from the UK HP test, containing information on the test candidates (e.g. their age and gender), as well as their performance on the HP test (e.g. the scores they received for each of the HP video clips they saw and their overall performance on the HP test).

Before the data was analysed, the decision was made to remove data from test candidates who had attempted the HP test more than once. This decision was made in order to avoid any effect that multiple test attempts may have on test performance, allowing for a direct comparison between test candidates who were attempting the test for the first time. The result of this was that data from a total of 26,452 candidates was included in the multidimensional scaling. The data divided into 20 separate sets corresponding to the 20 different HP test sets used by the DVSA, with each HP test set containing 14 hazard videos. The data was then further divided between the test candidates that had passed the test and those that had failed the test. The scores that the test candidates received for each of the 14 HP videos were placed into an ALSCAL multidimensional scaling in order to explore if there are particular hazards (and hazard properties) that are best at discriminating between candidates who pass and fail the HP test.

After carrying out the multidimensional scaling it became clear that the dimensions of the scaling solutions were being heavily skewed by the presence of the double hazard video contained in each HP test set. The decision was made to remove the double hazard video, resulting in sets that contained 13 HP videos. The multidimensional scaling was then carried out again. Once the scaling solutions had been created, a hazard classification guide created by the DVSA was used to assign the individual hazards to specific categories based on the nature of the hazards and the environment in which they took place.

After examining these newly classified hazards in each of the 20 HP tests sets (for candidates who both passed and failed the test) it became clear that there was no clear underlying structure in the data indicating that certain hazards, hazard types or tests sets were better at discriminating between those who passed and those who failed the HP test. While there was some evidence of an underlying structure relating to certain types of hazards and environments (e.g. hazards that involved erratic driving) this was not consistent across the majority of the tests sets, with other scaling solutions giving predominance to hazards caused by obstacles in the road. These inconsistencies also applied to the environments that the hazards took place in. There were even inconsistencies within a number of scaling solutions with hazards that shared the same cause and environment placed in very different locations within the scaling solutions. Ultimately, the different hazard sets appeared to be quite idiosyncratic with regards to their scaling solutions.

One explanation for these results is that the UK HP test is very well balanced across the individual hazard videos and the tests sets. It is also possible that in order to uncover underlying structures in the HP test data, a slightly different approach may be needed. For example, it may be necessary to focus on the precursors to the hazards instead of just looking at the cause of the hazard itself. By taking into account these more subtle elements of the hazard videos, as well as looking at the main cause and environment the hazards take place in; any underlying structures in the hazard data may become more prominent.

Study 2 continued to explore possible avenues of enhancing the UK HP test by comparing standard HP videos (which were taken from the training set for the UK HP test and are representative of the HP videos used in the test) with simulated versions of those hazards that were created and presented in a dedicated driving simulator. One of the major issues facing video-based HP tests is whether or not they are a reliable and valid test of HP ability as the method is far removed from actually driving a car on the road and responding to hazards that pose a tangible risk.

Study 2 exploited the advantages of simulation by testing HP in a dedicated driving simulator and comparing that to a standard video-based HP test. The simulator itself consists of a 180 degrees field of view projection screen and a stripped-down driving rig (a seat, a steering wheel and a gear lever integrated together). Three projectors are used to present the simulated driving scenarios onto the projection screen. Eight simulated driving scenarios, based upon eight videos from the DVSA HP training set were created and then compared with those HP videos. For both the simulated and video hazards, participants sat in the driving simulator. The simulated hazards were projected onto the 180-degree screen while the videos were presented on the television screen placed in front of the driving rig. For both the simulated and video hazards the participants were required to press a steering wheel mounted button when they perceived a driving hazard. The simulated hazards also required the participants to use the driving rig to drive through the simulated scenarios that contained the hazards. Their task was to react to and avoid having a collision with the hazards. Participants completed two HP tests (either video or simulated) and were placed into one of four groups: video-video, video-simulated, simulated-video or simulated-simulated. Behavioural responses (hazard scores and the number of button responses) and physiological responses (skin conductance and heart rate) were compared between the video and simulated hazards.

The results revealed significant differences in behavioural responses (button responses and hazard scores) and psychophysiological responses (skin conductance) between the videos and simulations. The simulated hazard scenarios were associated with significantly fewer button responses and lower HP scores, as well as increased skin conductance levels in comparison to the hazard videos. Heart rate was not found to be significantly different between the videos and simulations however. These results indicate that driving simulator based HP testing evokes significantly different responses from drivers compared to HP tests in a video-based setting (much like that used in the UK HP test). This finding questions how representative video-based HP tests are in relation to the process of detecting and responding to hazards while driving on the road. There was also some evidence of positive HP score transfer, with simulated hazards aiding subsequent performance on video hazards. This finding indicates that simulated hazards could help provide training for learner drivers to improve their HP performance. However, it is possible that this result may have been a by-product of the design of the study.

There are some alternate explanations for the above results including the dual-task issue associated with the simulated hazards impacting button responses and hazard scores, as well as the significant visual difference between the video and simulated hazards which may have had an effect on the skin conductance levels.

The next step in the thesis was to begin investigating what role the frontal lobes play in driving using fNIR. Study 3 focused on validating the fNIR device used in this thesis. In order to investigate what role the frontal lobes play in driving, fNIR was used to monitor frontal lobe activity across a range of different driving tasks. This process began in Study 3 by first validating the fNIR device with three classical frontal tasks: the WCST, the CBT and the COWAT. Two WCST were utilised in the study, a standard WCST (task A) and a WCST in which the participants were informed of the sorting criteria every four turns (task B). The baseline task consisted of a WCST in which the stimulus card exactly matched one of the target cards in every dimension. Two different block tapping tasks were also used, a standard CBT as well as a block suppression task (BST) in which distractor blocks were illuminated at the same time as the target block in order to provide a task distraction. The baseline task consisted of a CBT in which the target block remained the same in each trial. For the COWAT, the participants were tasked with producing as many words as they could think of beginning with a certain letter within the space of 1 minute. Four different letters were used ('F', 'A', 'R' and 'S'). These generation tasks were compared with baseline tasks in which the participants were required to read aloud four sets of words that began with the letters used in the generation task. fNIR was used to monitor activity during these tasks, as well as during carefully designed control conditions (baseline tasks) for each of the tasks.

The results of the WCST revealed that the standard WCST (task A) resulted in significantly higher levels of frontal lobe activity compared to baseline. The modified WCST (task B) did not result in a significant increase in frontal lobe activity compared to baseline though. These results were in line with prediction as task A was designed to be the most difficult and demanding task, while task B was only slightly more demanding than the baseline task. The failure to find a difference in frontal lobe activity between tasks A and B is in line with previous research from Lie et al. (2006) who used the same tasks and also failed to find a difference between them using fMRI. The WCST findings are also in line with other neuroimaging and lesion studies that have demonstrated that frontal lobes are vitally important for the WCST (Haines, 1994; Milner, 1963; Rezai et al., 1993).

Meanwhile the word generation task of the COWAT was found to results in significantly higher levels of blood oxygenation than the reading task. This finding was also in line with expectation as the generation task is significantly more demanding that the reading task. The results of the COWAT support previous neuroimaging studies that have found that the COWAT significantly increases activity in the frontal lobes (Warkentin and Passant, 1997; Frith, 1995), as well as lesion studies that have shown that damage to the frontal lobes has a significant detrimental impact on COWAT performance (Baldo et al., 2001; Butler et al., 1993; Stuss et al., 1998).

The CBT and the BST however did not result in a significant increase in frontal lobe activity compared to the baseline task. No significant difference in frontal lobe activity between the CBT and the BST was found either. These results were not in line with the prediction that the BST and the CBT would result in significantly higher levels of frontal lobe activity than the baseline task, as well as the prediction that the BST would results in greater frontal lobe activity than the CBT. This particular experiment was designed on the basis of the Toepper et al. (2010) design, with a difficulty hierarchy of BST>CBT>baseline, so the failure to find significant differences is surprising given that Toepper et al. (2010) did manage to find significant activation of the frontal lobes during the CBT and BST. The fact that the behavioural results from the CBT demonstrated that the participants found the BST the most difficult, followed by the CBT and then the baseline task made the fNIR results more surprising. Other neuroimaging studies (Bor et al., 2001; Owen et al., 1996) and lesion studies have also established that the frontal lobes are utilised during the CBT (van Asselen et al., 2006; Bor et al., 2006; Teixeira Ferreira et al., 1998). Interestingly, further examination of the CBT data revealed that, although not significant, there was a trend towards increased blood oxygenation in the BST compared to the CBT and the baseline task.

There are a few possible explanations for these findings. The CBT is largely a test of VSWM, the functions of which are associated with a wide network of neural areas, including many areas outside of the frontal lobes such as the parietal lobes (van Asselen et al., 2006; Toepper et al., 2010). It is therefore possible that the failure to find differences in frontal lobe activity between the

CBT, BST and the baseline task was in part due to the CBT and BST primarily activating areas outside of the frontal lobes such as the parietal cortex that the fNIR device is not able to record from. It is also important to mention that previous neuroimaging studies that have studied the CBT have used techniques such as fMRI and PET, which could also go some way to explain the discrepancy in results.

A further explanation for the results may come from the BST findings. The fact that the BST results revealed a trend towards increased blood oxygenation compared to baseline could provide a further explanation. The major difference between the BST and the CBT is the inhibitory aspect contained within the BST as the participant must supress and inhibit the spatial information of the distractor block, and as mentioned previously the DLPFC is associated with inhibitory processes (Weinberger et al., 2006). Toepper et al. (2010) reported that the BST and CBT differed in the brain areas activated, with the BST activating the DLPFC, whereas the CBT activated the VLPFC as well as the DLPFC. The fNIR device used in this thesis is best suited to recording activity from the DLPFC, leading to the possibility that the CBT resulted in activation being spread out across the VLPFC (which the fNIR device may not have been able to record activity from) and the DLPFC (in which the activity may not have been large enough). The trend towards increased activity in the BST may have been due to the inhibitory elements of this task activating the DLPFC, with failure to find a significant increase in frontal lobe activity for the BST due to the inhibitory processes in the task not being great enough to result in significant increases in DLPFC activity.

The conclusion taken from these results was that the significant increases in blood oxygenation in the WCST and COWAT were due to the effect of task workload and inhibition. The fact that the BST, a task that evokes inhibition as well as increased task workload due to the increased difficulty of the task, displayed a trend towards increased blood oxygenation compared to the CBT lends further support to this idea.

Based upon the findings of Study 3, a set of simulated driving tasks that elicited varying levels of workload and inhibition were designed in order to investigate if the frontal lobes play an important role in driving. Given the major functions of the frontal lobes, it appears logical that the frontal lobes will be involved in the process of driving; and research has demonstrated that workload and inhibition may be important factors in driving. Previous driving research, including a number of studies that used fNIR, has indicated that frontal lobe activity during driving is heavily linked to workload (Kojima et al., 2006; Tsunashima and Yanagisawa, 2009; Yoshino, Oka, Yamamoto et al., 2013a, 2013b). Research has also shown that inhibition plays an important role in driving, namely with regards to controlling impulsive and potentially dangerous driving behaviours (Cheng and Lee, 2012; Fillmore, Blackburn and Harrison, 2008; O'Brien and Gormley 2013).

Study 4 consisted of a set of simulated overtaking and following tasks that were designed to vary the levels of workload and inhibition required. Overtaking and following tasks were chosen as they allowed for a straightforward manipulation of task workload and inhibition. The simulated overtaking task placed the participants on a 3-mile stretch of single carriageway road populated by vehicles traveling in both directions. The participants were instructed to overtake as many vehicles as they could, while maintaining a safe driving style and not crashing. The following task took place on the same single carriageway road, only this time the participants were simply instructed to follow the vehicle in front. These two tasks were the 'active' tasks. Two further 'passive' driving tasks were created, which consisted of watching replays of the overtaking and following tasks. The passive tasks were introduced as they allowed a further level of task workload and inhibition manipulation.

The logic behind the design of study 4 was that the overtaking task would require inhibition (namely from inhibiting unsafe overtakes that could lead to a collision) and workload, while the following task would contain significant workload (from the need to operate the simulator controls to drive through the simulated route) but with the inhibitory element removed. The passive tasks were designed to remove the workload element of the tasks entirely while keeping the visual stimuli identical. As well as recording frontal lobe activity using fNIR, subjective levels of task workload were recorded using the NASA-TLX workload scale in order to track the participants perceived level of workload across the four driving tasks.

The results of the study revealed a strong correspondence between the levels of frontal lobe activity and the subjective workload scores. As expected, frontal lobe activity was higher during the active tasks than in the passive tasks, as well as being higher in the active overtaking task than in the active following task. The subjective workload scores mirrored this, with higher levels of perceived workload during the active tasks compared to the passive tasks and during the overtaking tasks compared to the following tasks. Overall the results of study 4 suggest that frontal lobe activity was primarily being driven by task workload without any clearly separable contribution from inhibition required during the tasks.

A consistent finding in driving research is that age and driving experience can have a significant impact on driving behaviour. This includes differences in HP ability, crash and fatality rates, road scanning and general risk perception (which relates to inhibitory control and impulsiveness). It is possible that these differences are in part related to maturational differences in the frontal lobes. Significant differences in driving behaviour as a function of gender have also been extensively reported, with male drivers being more prone to risk-taking behaviours while driving.

Study 5 was designed to explore what effect driver age, experience and gender had on frontal lobe activity and driving performance. Based upon the findings of study 4 the design was altered in a few important ways. A potential issue with the design of study 4 was that the active following task and the passive tasks were too easy in comparison to the active overtaking task. As a result of this the passive tasks were removed and the following task was carried out alongside four overtaking tasks that differed in traffic density (i.e. how many vehicles were present in the driving simulations). At the lowest traffic density overtaking opportunities were relatively frequent, whereas in the highest density there were only limited opportunities to overtake. This means that the overtaking tasks should have differed in the levels of inhibition they required (as the participants would have needed to spend more time inhibiting unsafe overtaking manoeuvres as the traffic density increased) while task workload remained roughly equal. This change was made in order to avoid having such a large difference in both task inhibition and workload between the following and overtaking task that was present in study 4, as the different traffic densities would result in a more gradual increase in task inhibition relative to the following task without a necessary increase in workload. In turn this would potentially allow for a disentangling of the relative effects of workload and inhibition on frontal lobe activity. A further change involved adding a subjective inhibition measure to the NASA-TLX scale in order to better gauge the levels of perceived inhibition present during the tasks, which could in turn be used as a reference for the fNIR results.

Participants were recruited so that drivers of different ages, driving experiences and genders were represented. Depending on their age the participants were assigned to either the 'young' or 'old' groups, while their driving experience determined if they were part of the 'novice' or 'experienced' group. Male and females were equally represented in these different groups. Frontal lobe activity, perceived workload and perceived inhibition were recorded across all five driving tasks.

As expected, the perceived workload and inhibition scores displayed significant differences between the overtaking tasks and the following task, with the overtaking tasks reported to be more demanding and requiring more inhibitory control than the following task. Females and inexperienced drivers also reported higher levels of perceived workload across a number of the workload measures than males and experienced drivers. The perceived level of inhibitory control was also found to be higher for the overtaking tasks compared to the following task. Experienced and older drivers also reported higher levels of inhibitory control in comparison to inexperienced and younger drivers, for both the following and overtaking tasks. Overall, these results indicate that the overtaking and following tasks differed in task workload and inhibition as they were designed to.

When comparing levels of perceived workload and inhibition between the different overtaking traffic densities, the results were mixed. Perceived Inhibitory control was reported to have increased in a linear fashion across the four overtaking traffic densities. However, the only workload differences found across the four overtaking densities were for a reduction in perceived performance and higher temporal demand for novice drivers. These findings demonstrate that the task design was successful as the four overtaking densities differed in the amount of inhibitory control required but not in terms of task workload.

The number of successful overtaking manoeuvres completed was affected by traffic density, with a general reduction in the number of overtakes completed as the traffic density increased. This result demonstrated that increasing the traffic density had the desired effect of making overtaking more difficult. Interestingly, there were differences in the number of overtakes completed by males and females across the different overtaking densities, with males tending to overtake more than females. This finding supports that idea that male drivers tend to take more risks than female drivers.

As predicted, frontal lobe activity was found to be significantly higher in the overtaking tasks compared to the following task, with the results showing a good level of correspondence with the subjective workload and inhibition scores. However, no significant differences in frontal lobe activity as a function of age, gender or driving experience were found when comparing overtaking with following, despite the fact that the subjective workload and inhibition scores differed among these groups. No significant difference in frontal lobe activity was found between the different overtaking densities, which largely mirrored the subjective workload scores but was not in line with the inhibitory control scores that increased across the traffic densities. No effect of age, driving experience or gender on frontal lobe activity was found either, which was largely in line with the subjective workload and inhibition scores. Overall these results indicate that task workload is the primary factor driving frontal lobe activity across the driving tasks as the overtaking tasks were designed to gradually increase inhibitory control while maintaining task workload. The fact that the inhibitory control scores increased in a liner manner across the increasing traffic densities without a corresponding increase in frontal lobe activity helps demonstrate this.

A limitation of study 5, and a possible explanation of the failure to find differences in frontal lobe activity as a function of age and driving experience, relates to the sample of participants. The range of ages and driving experience in study 5 was potentially too limited. In this study the ‘young’ driver group had an average age of 19 years. It is possible that, in order to observe differences in frontal lobe activity, younger drivers who are 17 years of age (the age at which driving becomes legal in the UK) may need to be tested as well as those as old as 25. 17 year-olds have been found to have the highest fatality and collision rates of all drivers (ECMT, 2006; Forsyth, 1992) while research from the likes of Sowell et al. (1999b) has made it clear that the frontal lobes do not reach full maturation until about 25 years of age, which is also the age at which fatality rates among drivers begin to fall significantly (ECMT, 2006). It also appears that the effects of age and experience on driving are not linear and have a greater effect the younger drivers are, with this effect having even been seen between 16 and 17 year old drivers (Maryland et al. 2001). This suggests that this youngest group of drivers may be very important when it comes to finding differences as a function of age. Jongen et al. (2011) found that that inhibitory control was lower in the 17-18 year olds than 22-24 year olds, showing that there are differences in the executive functions in people of this age. This reduced inhibitory control was accompanied by increased risk taking. This may explain why drivers in their late teens demonstrate an even greater propensity to be involved in accidents than young adults. The interactions between age and driving experience on collision and fatality rates could also mean that not having a wide enough age group of drivers may have also masked some of the effects of experience.

Study 6 built upon all of the previous studies in the thesis by combining the two major research strands (HP testing and the frontal lobes and driving) by investigating HP with the use of fNIR to monitor frontal lobe activity. The major difference in this study compared to study 2 was the comparison of simulated hazards with replays of simulated hazards, as opposed to video hazards. This change was made for a number of reasons. The first reason was to make the visual presentation of the hazards more similar across conditions in order to remove any effects that visual differences (i.e. the difference between presenting video hazards on a T.V. compared to presenting simulated hazards on a 180 degree projection screen) in the presentation might have had on the results. Comparing simulated hazards with replays of those hazards also removes any possible effect that differences in the content of the HP videos and simulations may have. A further reason is that the replay tasks provide a means of eliminating the dual-task issue of the simulated drives while providing the same visual stimulus. Lastly, replay hazards provide a potential means of combining simulated elements into the HP tests, allowing them to better represent real on-road HP as they are essentially the intermediate between the video-based hazards and the simulated hazards. As in study 2, button presses and hazard scores for each of the hazards were recorded.

The behavioural results from study 6 were in line with those of study 2, with simulated hazards associated with significantly fewer button responses and lower hazard scores than the replay hazards. One possible explanation for the results is that the simulated HP task was a more difficult and demanding, with participants struggling to detect and respond to the hazards compared to those in the replay task. An alternate explanation is that the results are due to the dual-task present in the simulated task (i.e. the participants had to use the driving controls to navigate through the simulated world as well as responding to any hazards with a button press). It is possible that the participants in the simulated task prioritised the driving controls over button presses whereas those in the replay task did not have this issue, as they were not required to drive through the simulated environments. There were also significant differences in the hazard scores and button presses between the eight hazards, with some hazards receiving more button presses and higher hazard scores than others. This again demonstrates that the button press methodology is sensitive to differences in content between the hazards.

Contrary to expectations, no significant difference in frontal lobe activity was found between the simulated and replay hazards. Given that studies 3, 4 and 5 indicated that frontal lobe activity recorded by the fNIR device was primarily associated with task workload, one interpretation of these findings is that the workload in the simulated driving task was primarily due to the HP process itself, as opposed to operating the driving controls and driving through the simulation while avoiding any hazards (and any extra difficulty that this introduced into the task compared to the replay task). This would explain the lack of any difference in frontal lobe activity compared to the replay task, as the process of detecting hazards was also present in the replay task. Differences in frontal lobe activity were found between the different hazards however, indicating fNIR is sensitive to these differences. Overall, the findings from study 6 would suggest that replay-based HP tests are able to evoke similar levels of workload and cognitive processing as simulated driving, suggesting that replay-based HP could provide a good way of implementing simulation into the HP test in order to better evoke elements of on-road HP.

9.3 Conclusions

9.3.1 Hazard Perception Testing

With over 1 million people killed on the roads around the world in 2013 (WHO, 2013), it is clear that road safety is an issue of major importance. A startling finding from dozens of countries around the world is the significant overrepresentation of young novice drivers in road crash and fatality statistics.

Globally, 16-24 year olds represent 27% of all drivers killed on the roads despite only making up 10% of the population (ECMT, 2006). The death rate for 18-24 year olds is roughly double that of older drivers, while the single biggest killer of 15-24 year olds in OECD countries is motor vehicle crashes (ECMT, 2006).

One of the possible reasons for the high representation of young novice drivers in crash and fatality statistics is that these drivers have limited and underdeveloped HP abilities in comparison to older and more experienced drivers. A number of studies have demonstrated a link between slow HP reaction times and an increased likelihood of a driver being involved in crashes (McKenna and Horswill, 1999; Pelz and Krupat, 1974; Quimby et al., 1986;), while Drummond (2000) linked slow HP reactions times to an increased likelihood of being involved in a fatal crash within the first year of driving. Importantly, research has found that age and experience affect HP reaction times, with young drivers having slower HP reaction times than middle-aged drivers (Quimby and Watts, 1981), while inexperienced and untrained drivers have slower HP reaction times than experienced and trained drivers (McKenna and Crick, 1991; McKenna and Crick, 1994; Wallis and Horswill, 2007). The link between longer reaction times and higher crash rates has also been found when controlling for age, driving experience and simple reaction times (Quimby et al., 1984). As detailed in chapter 1, a number of explanations have been put forward to explain these differences in HP, including differences in how young novice drivers scan the road (Chapman and Underwood, 1998; Mourant and Rockwell, 1972; Pradhan et al., 2005), how they utilize precursors to hazards (Crundall et al., 2012) and how they perceive risk while driving (Wallis and Horswill, 2007).

Based upon these findings researchers began investigating HP testing, finding that it was HP was a trainable skill (McKenna and Crick, 1994; Mills et al. 1998), that testing could discriminate on the basis of driver experience and potentially even accident liability (Grayson and Sexton, 2002). As a result of these findings on HP testing, governments in a number of countries have started to implement, or are considering implementing HP tests into their pre-existing driver testing programmes. In the UK, HP testing was implemented into the driving test in 2003 with the aim of improving road safety by reducing the number of fatal and non-fatal collisions that newly qualified drivers are involved in. Since the introduction of the HP there had not been a systematic study of HP data. Wells et al. (2008) did investigate what effect the introduction of the HP had on the driving of newly qualified drivers. However, beyond this study there has not been a large amount of research into the UK HP test since its launch. As a result of this, one of the major aims of this thesis was to provide new research into the HP test, in order to better understand what factors affect whether or not people pass the test. This in turn can help inform the DVSA on possible ways to improve the HP to better discriminate between drivers with and without the necessary HP abilities.

HP research has largely focused on three methods of testing: video, simulated and on-road. Although simulated and on-road HP testing are increasingly being investigated through research, video-based methods are still the predominant method used in research and real-world testing of HP performance (e.g. the UK HP test).

HP testing has primarily involved the use of video hazards for a number of reasons. One of the major reasons is that video-based methods provide a high level of experimental control, with each participant guaranteed to view exactly the same content and having the opportunity to respond to exactly the same hazards in a given video. This can be very difficult to achieve with simulator or on-road based HP methods. For example, people who drive slightly too fast in a simulator HP test may miss certain hazard precursors or cues as they have driven passed a certain point before the hazard had time to trigger. As a result of this they will not respond to the hazard and therefore receive no score for that hazard. For the analysis this becomes problematic, as this score would not actually be representative of their HP ability, rather it would represent a problem with the methodology. HP testing on public roads is next to impossible as there is no way to control the environment, meaning that there would be no way of having each test candidate interact with the same hazards in the same way. Having people drive on test courses with staged hazards overcomes this issue, however the same issues of driver behaviour that affect simulated HP testing (e.g. driver speed) would cause problems. On-road HP testing, even that taking place on closed roads, also introduces a number of serious safety and ethical concerns.

Economic reasons are also an important factor in the HP testing method used. Each year over one million candidates take the HP test across hundreds of test centres in the UK. The video-based method used in the HP test allows for a large number of people to be tested quickly and efficiently as each test centre can host a large number of computers on which the test is carried out on, with only a few member of staff needed to oversee the operation. Test results are calculated and given out Building multiple driving simulators in each of these test centres is not feasible from an economic or practical standpoint. On-road HP testing would also carry increased costs, as a large number of driving testers would need to be trained and employed.

Although there are a number of practical and economic drawbacks to simulated HP tests, there are a number of important advantages that make this method attractive. The biggest advantage of staging HP tests in a high-fidelity driving simulator is the level of realism it provides in comparison with traditional video-based methods. Driving simulators allow for the imitation of real driving without the practical, safety and ethical issues associated with testing HP on the road. However, even though driving simulator fidelity is improving each year, the visual fidelity of simulation cannot match that of a high quality hazard video. Driving simulators also allow for multiple response methods. The button method used in this thesis can be used but the use of car controls (e.g. braking, steering) is also be possible. Given that safely responding to hazards while driving is often dependent on a proper response using car controls, scoring the performance of a test candidate on the basis of control inputs could make more sense than simply using a button press to record a response time.

A further advantage simulated HP has is the ability to quickly and easily change the content of hazards. For example, new vehicles can be added to reflect the introduction of new models into the market. Environmental conditions such as the time of day and weather can also be changed very easily. One of the biggest drawbacks of the video-based method is that is that the only way to change the content of hazard videos is to completely re-film them. Not only is this time consuming and expensive but there is no guarantee that it will be possible to desired hazard. As a result of this, if a hazard becomes out-dated and needs to be filmed again, it may simply have to be replaced with an entirely different hazard. One of the major issues the DVSA has dealt with recently is the fact that the hazard videos became dated and the visual quality was not up to modern standards. Instead of re-filming the hazards the DVSA instead decided to create animated versions of the original video hazards.

Due to the economic and practical constraints of simulated HP testing but the clear advantages it can provide, one of the major aims of this thesis was to provide the DVSA with ways of improving the current HP test without undertaking a radical overall of the test. In study 2, simulated driving was used as an analogue for on-road driving due to the experimental and ethical issues associated with testing HP on public roads. Although simulated driving cannot exactly replicate the physiological and behavioural aspects of driving on public roads, research has shown that it can evoke similar responses (Konstantopoulos et al., 2010; Shechtman et al., 2009). Therefore, the implication of these findings is that video-based HP tests may not evoke the psychophysiological and behavioural responses associated with driving on the roads. This in turn suggests that introducing elements of driving simulation into HP tests could potentially lead to a more realistic and improved method of assessing the HP skills of drivers.

Study 6 further supports the idea that introducing elements of driving simulation may improve the hazard perception test. The fNIR results from study 6 revealed no significant differences in frontal lobe activity between the replay and simulated hazards, indicating that task workload was roughly equal across the two tasks. Given the differences between video hazards and simulated hazards found in study 2, the results of study 6 may indicate that the replay hazards were able to closely mirror the HP processes present in the simulated hazards. Given that the simulated hazards are designed to be an analogue of on-road driving, these results could suggest that the replay hazards provide a means of bridging the gap between video hazards and real world driving hazards without the need for a full implementation of driving simulations.

However, as mentioned above, there are a number of important practical and economic reasons that the UK HP test employs the video-based method. These issues mean that until simulation technology becomes more affordable and practical, any changes to HP testing must be made within a certain set of parameters, with any changes made to the HP being limited on the basis of cost and practicality. It is therefore important to come up with ways of introducing elements from simulated driving into the HP test, which can be achieved taking these constraints into account.

The early HP tests created by Watts and Quimby (1979) and Quimby and Watts (1981) attempted to make HP more like on-road driving by having participants sit inside of the body of a car. However, McKenna and Crick (1991) argued that introducing elements into a HP test that aim to increase the similarity to on-road driving results yet are not used results in ‘redundant realism’, with the most important aspect of the test being the hazard videos themselves. If simply having test candidates sit in a driving rig while watching hazard videos is not sufficient, introducing replay hazards in an environment more similar to that of driving (e.g. by using high fidelity and large filed of view projection screens) may produce an experience more representative of HP while driving on the roads without needing to spend money on multiple car rigs.

One of the major difficulties in this thesis was creating simulated hazards that accurately matched the video hazards. While it was possible to closely mimic the behaviour and placement of vehicles and pedestrians in the HP videos, there were inevitably some differences. One of the reasons for this was that the simulator’s AI was not able to perfectly match the behaviour of some vehicles and pedestrians. A further important factor was the limitation in accurately matching the driving environments themselves. The driving simulator used in this thesis has a predefined simulated ‘world’ in which simulated driving scenarios can be created. The problem this creates is that any hazards created are restricted by the properties of the simulated world itself but the actual process of creating more simulated environments can be very time consuming and expensive. As a result, it was not possible to create exact replicas of the road environments present in the HP videos. This in turn has a knock on effect for the vehicles and pedestrians. This creates the possibility that response differences to the video and simulated hazards may have in part been due to differences in the content of the videos and simulations.

Research has shown that differences in the content of HP clips can impact responses. Chapman and Crundall (2011) compared videos from the UK HP test set (the same videos used to create the hazards in this thesis) with animated versions of those videos in order to prepare for the introduction of animated clips into the HP test. Their findings revealed that even relatively small differences in the content of the video and animated clips had significant effects on button responses. Comparing the button responses between the video and animated versions revealed that although the overall pattern of responses was similar, there were some significant differences. Differences in the response rates, peak response rates and button response timings were found.

The authors found that even small differences in the content of the video and animated hazard clips could produce these differences. These differences included there being a different numbers of pedestrians or parked cars in an animation. For example, an extra pedestrian in one of the animated clips may have resulted in the learner drivers attending a different part of the visual scene, altering their responses compared to the standard video version. Differences that occurred near and far from the scoring hazard both affected responses. For example, a white van becoming visible slightly later in one of the animated hazard videos resulted in a significant difference in button responses compared to the standard video version of that hazard.

These findings demonstrate that if simulated hazards were to be implemented into the UK HP test they would need to match the hazards currently used exactly, as any small changes could significantly affect responses. It is also important to note that creating simulated driving environments can be time consuming and expensive. For example, in order to create driving environments that exactly match the roads present in the UK HP test, a method such as collecting LIDAR scans of each of the road environments would be needed. Road environments that are representative of but not identical to those used in the UK HP test videos could be created without resorting to a method such as LIDAR, however this would then run the risk of differences being present in the hazards which in turn will affect responses. Regardless of the chosen method, each new simulated hazard would need to be re-validated to determine if they are able to discriminate between experienced and inexperienced drivers (which is the fundamental basis of the UK HP test). These findings demonstrate the technical difficulties involved in creating simulated hazards. It should however be noted that the similar ordering of the hazard scores for the simulated and video hazards demonstrates that the simulated hazards were largely successful in imitating the video hazards.

The potential usage of simulated hazards raises an important issue, what method of scoring HP performance should be utilised? A significant limitation of the button-press method used in this thesis is that by requiring both a button response and car control in the simulated hazard conditions we created a dual task situation that is not required in traditional HP testing, or in actual driving. Most research looking at hazard responses in a simulator has avoided this issue by measuring hazard perception responses through the use of simulator control actions such as braking and steering. For example, using the application of the brake or a change in steering angle to determine when a participant has detected a hazard. The difficulty with this method is that not only would the timing and scoring of such measures would be very hard to achieve in itself (e.g. determining exactly which responses which counted as an acceptable detection of the hazard) but this method would be completely different to that of button pressing in a traditional HP test, meaning that direct comparisons between environments in terms of HP scores would not be possible.

It would be interesting to have participants drive these simulated HP scenarios without button pressing. It is likely that it would be very difficult to create comparable scores based on purely driving measures, but it would be interesting to explore the degree to which driving behaviour itself was changed by the secondary button-pressing task. This method also provides a possible alternative for testing HP in the UK HP test if simulated elements were to be introduced into the test. Given the large difference between pressing a button to respond to a hazard during a test and the reacting to a hazard to avoid a collision during driving, implementing a testing method that scores performance based on driving inputs could provide some increased efficacy into HP testing. As well as being used for simulated hazards, this method could also be used in conjunction with the video hazards (or in the case of the UK HP test the animated clips that have been introduced). For example, test candidates could sit and watch hazard videos while sitting in a simple driving rig and respond to any hazards with a brake press or a change in steering angle for example. However, such a method would require a full revalidation of all hazards. The UK HP test was validated on the basis of discriminating between inexperienced and experienced using button presses. Changing the response method would require this method to be fully validated with each hazard used in the test.

The issue of HP testing is complex and its scope goes far beyond that of any research laboratory, encompassing important public policy aspects and a number of complex practical and economic issues. Ultimately more research is needed to establish what is the best future direction for HP testing. However, from the research carried out in this thesis it does appear that implementing some elements of driving simulation could help improve HP testing. This could take the form of having a visual environment more similar to that found while driving or possibly utilising a different method for scoring HP performance, namely using driving control inputs. Changes such as these may be able to make HP testing more representative of real world driving. Future improvements and cost reductions for high-fidelity driving simulators may ultimately allow for the introduction of more simulated elements into the HP test.

9.3.2 fNIR and Driving

The use of fNIR as a means of recording frontal lobe activity during driving was explored in this thesis as it has the potential to be a relatively cheap, easy to use, portable and flexible method of studying frontal lobe activity in a wide range of driving environments such as dedicated driving simulators and in real cars on the road. Methods such as EEG and MEG are ruled out due to the fact that the electrical fields generated by a driving simulator would interfere with the recordings, while these methods are not portable enough to be used in real cars on the road. Traditionally, research investigating brain activity during driving has utilised methods such as fMRI. A number of studies have been carried out in which the researchers have set up low-fidelity driving simulators in an MRI scanning room (e.g. Graydon et al., 2004; Horikawa et al., 2005; Calhoun and Pearlson, 2012).

One problem with using fMRI to study driving is that the scanning room environment limits the use of realistic driving controls such as pedals, a steering wheel and a gear lever (Horikawa et al. (2005) used a joystick control while Graydon et al. (2004) used button controls). Even if more realistic driving controls are used other problems remain. Using fMRI requires the participants to lay down on their backs while engaging in a driving task and also results in a restricted field of view (although goggles can be used to improve the field of view). fMRI is also very sensitive to movement, which places restrictions on the ability to simulate driving in as realistic manner as possible as driving in a car naturally involves a certain amount of movement, including significant head movement. So while fMRI does offer high levels of detail regarding which areas of the brain are activated during driving tasks, it is much more restrictive in terms of what driving research can be undertaken as well as being more expensive and time consuming than fNIR.

Continuous wave fNIR devices like the one used in this thesis are relatively affordable, easy to set up and use and also have the advantage of being small and portable. This means that they can be used in a wide variety of driving environments, including driving simulators and in cars. For example, in this thesis the use of fNIR allowed for frontal lobe activity to be recorded while participants drove in a dedicated driving simulator with car controls and a 180-degree field of view projection screen (which can be also provide a view of what is taking place behind the driver through the use of a TV screen). Although fNIR is sensitive to motion artefacts, subjects are still able to freely move their head and other body parts. This freedom provides researchers with the ability to monitor brain activity across a wide range of driving tasks and driving environments. This thesis used fNIR in a dedicated driving simulator with HP and overtaking tasks, while other researchers have used fNIR in driving simulators and on public roads to test a number of driving tasks such as following with and without adaptive cruise control in a driving simulator (Tsunashima and Yanagisawa 2009), as well as more general driving tasks and manoeuvres such accelerating, breaking and reversing while on public roads (Yoshino, Oka, Yamamoto et al., 2013a, 2013b). fNIR has even been used to study simulated train driving (Kojima, Tsunashima and Shiozawa, 2005) and motorbike riding on a racing track (Kawashima, Matsumoto and Tanimoto, 2014).

fNIR devices also have the advantage of having very high sampling rates and temporal resolutions, with some devices having sampling rates as high as 100Hz, although the device we used in this thesis was only sampling each region at a rate of 1Hz. In comparison, fMRI can have temporal resolutions in the region of several seconds (in the case of high spatial resolution scans). A weakness shared by both fNIR and fMRI is that that they are both dependent on the haemodynamic response, which is an indirect measure of neural activity and lags behind the underlying neuronal events by a few seconds (Logothetis et al., 2001). This ultimately limits the temporal resolution of both methods.

However, compared to other neuroimaging methods used for driving research, fNIR does have a number of drawbacks. One of these drawbacks is that fNIR is only able access and record neural activity from regions 1-2cm deep, limiting the areas of the brain that can be studied. A further issue is that fNIR does not have the spatial resolution that fMRI has in order to accurately localise neural activity. fNIR devices have spatial resolutions ranging from 1 to 10cm, whereas the temporal resolution fMRI can be as high as a few millimetres. Although methods exist to help localise neural activity recorded with fNIR, the ability to accurately determine which specific areas neural activity is associated with is still limited in comparison. For example, software exists that can produce activity maps that provide a visual representation of changes in haemodynamic activity across the different fNIR channels. These activity maps can then be overlaid onto a structural MRI scan to provide a general approximation of the areas of the brain that are being activated during a task.

Ultimately, the combination of limited depth and spatial resolution means that even this method is limited in comparison to methods such as fMRI. These drawbacks mean that fNIR is not suited to driving studies that want to record activity from areas deeper within the brain and/or need high levels of spatial resolution. The particular fNIR device used must also be taken into account. The 16-channel device used in this thesis only provides coverage of frontal areas, namely the DLPFC. This means that any neural activity that is taking place outside of these frontal areas will not be recorded. As a result, if researchers are interested in recording activity from more areas of the brain or areas other than the frontal lobes, they will need to invest in fNIR devices with more channels. For example, there are fNIR devices with up to 100 channels that cover most of the head, allowing for recordings to be taken from frontal, temporal, parietal and occipital areas of the brain.

One factor that has emerged from using fNIR in this thesis is that it primarily appears to be measuring task workload. In study 3, increased frontal lobe was associated with increased task difficulty and workload in the WCST and COWAT. In studies 4 and 5 increased frontal lobe activity was associated with the increased task workload of the overtaking tasks. Comparing the fNIR results to the subjective workload scores in studies 4 and 5 also supports this workload hypothesis, as there was a large correspondence between the subjective workload scores and the fNIR results. The failure to find increases in frontal lobe activity across the four overtaking densities, despite the linear increase in reported inhibitory control scores across the traffic densities (while workload scores remained mostly unchanged) further supports this idea.

This is supported by previous research that has also reported that fNIR seems to be primarily recording task workload. Tsunashima and Yanagisawa (2009) found that driving without cruise control resulted in more frontal lobe activity, likely because this involved greater task workload. Kojima et al. (2006) reported that frontal lobe activity increased when participants decelerated when driving in a train simulator. This effect was especially pronounced in the inexperienced train driver, which the authors attributed to increased task workload and task difficulty for the inexperienced driver. Yoshino, Oka, Yamamoto et al. (2013a, 2013b), used fNIR to record frontal lobe activity during a wide variety of on-road driving tasks. They reported increased activity in various areas of the frontal lobes, such as the PFC and frontal eye field, which was particularly prevalent during higher workload tasks such as accelerating and braking.

An increasing number of studies are actually employing fNIR (including the fNIR system used in this thesis) as a measure of task workload. There are a number of ways of measuring task workload including: subjective assessments (e.g. NASA-TLX), performance measures (e.g. accuracy and response time) and physiological measures (e.g. skin conductance, blood pressure, respiration). fNIR can provide researchers with a more direct and objective means of assessing mental workload during tasks, with the portability of fNIR allowing for use in a range of testing environments. So far fNIR has been used in a wide range of studies including measuring workload in air-traffic controllers (Ayaz et al., 2010), unmanned aerial vehicle pilots (Izzetoglu et al., 2014) and to explore the relation between metal workload and expertise (Bunce et al., 2011).

A further factor that has emerged is that in order to record significant differences in frontal lobe activity between conditions, there needs to be a large difference in the workload of the tasks. This is demonstrated in study 5, where despite there being significant differences in frontal lobe activity between the following and overtaking tasks (a result that was also found in study 4), no significant difference in frontal lobe activity was found between the four overtaking tasks with differing traffic densities. The fact that the subjective workload scores did not reveal many differences between the different overtaking tasks indicates that the failure to find differences in frontal lobe activity was due to the workload levels of the tasks being too close to one another. These findings indicate that is perhaps best not to use fNIR in studies where there are only small differences in the workload of tasks and only small differences in brain activity are expected. In such situations as this, a method such as fMRI would likely be a better choice. However, it should be noted that study 6 revealed that fNIR was sensitive to the differences between the individual hazards, with differences in frontal lobe activity found between the different hazards. These differences matched up with the differences in button presses and scores found between the different hazards. This finding suggests that fNIR can be sensitive to more general differences.

In summary, it appears as that fNIR is primarily measuring task workload and is therefore best used as a measure of task workload. fNIR is also probably best used when relatively large differences in frontal lobe activity are expected between tasks. Ultimately, despite the drawbacks of fNIR relating to its limited depth sensitivity and spatial resolution, fNIR provides a portable and flexible means through which to compare frontal lobe activity across a wide range of driving tasks and driving environments. The use of fNIR devices with full head coverage also provides a potential means through which to monitor a large number of areas of the brain during driving.

9.4 Further Research

While the research carried out in this thesis offers a solid foundation for better understanding the psychophysiology of driving as well as what factors affect HP performance, there is more research that can be carried out in order to get a deeper understanding of these subjects. One of the reasons for focusing on simulated driving in this thesis is the level of experimental control and safety it provides in comparison to driving on the public roads. However, in order to develop a deeper understanding of HP and the psychophysiology of driving it is necessary to study driving on the road. While methodological, safety and ethical concerns make HP testing on open public roads problematic, it would be possible to carry out versions of the studies in this thesis on closed roads with staged hazards. For example, participants could drive around a test course with hazard set up along it. They could be tasked with avoiding having collisions with those hazards. fNIR could also be used to monitor frontal lobe activity while performance could be measured using the button press method or by measuring a combination of driving inputs. Carrying out research that directly compares video, simulated and on-road HP would allow for a better understanding of the correspondence between the behavioural measures (hazard scores), physiological measures (skin conductance) and workload (fNIR) of HP in these different environments. It would also help provide further details regarding how well driving simulation can mimic real driving for other areas of driving research.

Although the HP typology failed to uncover any underlying structures in the HP test data, other methods may be able to discover whether or not there are particular hazards, types of hazards of clip sets that are best at discriminating between test candidates who pass and fail the test. One such method would be to classify hazards based upon the precursors to those hazards, as well as on the basis of the cause of the hazard and which environment it took place in. It is possible that an underlying structure in the HP test data may begin to emerge when these more subtle elements are taken into account when carrying out the multidimensional scaling.

As mentioned in Chapter 7, in order to further explore the effect that age and experience have on driving and whether the maturation of the frontal lobes is involved in these effects, it will be necessary to recruit drivers with a wider range of ages and driving experience. This could involve recruiting 17 year-olds who have a full driving licence as well as those that have only just started learning how to drive. Drivers at the age of 25 (the age of full frontal lobe maturation) with both large and small levels of driving experience could be recruited for comparison. Recruiting drivers over the age of 25, with high and low levels of driving experience, could also be used to provide a further point of reference in terms of the effect that age and driving experience have on frontal lobe activity and driving performance. This will allow for a more in-depth investigation regarding what effect frontal lobe maturation may have on driving as it would allow for a direct comparison between drivers with a much wider range of ages and driving experience.

One of the limitations of the driving simulations in this thesis was that there was a predefined simulated world in which they had to be created. Although as much effort as possible was taken to accurately simulate the HP videos in terms of vehicle and pedestrian location and behaviour, it was not possible to change the environments to match those in the HP videos. This ultimately affected the ability to accurately simulate the HP videos. Therefore, an extension to this research could involve creating a wider range of driving environments that are more able to mimic those present in the UK HP test. This would allow for a more accurate recreation of the video hazards, eliminating any effect that environmental differences (and the knock-on effects this has on vehicles and pedestrians) have on HP performance when comparing simulated to video hazards.

A drawback of the simulated driving tasks was the creation of a dual task that is not present in traditional video-based HP tests or in actual driving. During the simulated driving tasks the participants were tasked with responding to any hazards through a button response (similar to a video-based HP test) as well as driving through the simulated scenarios avoiding having any collisions (similar to on-road HP). This method was chosen as it allowed for the direct comparison. However, this does result in a method that is not the same as either video-based or on-road HP. This research could therefore be enhanced by utilising a different method of determining HP performance while driving in the simulator, in order to determine what affect the secondary button task has on driving behaviour. A combination of steering inputs, breaking and acceleration could be used assess any differences in driving behaviour. The current state of simulation does not make this easy, but the research in this thesis adds to the body of knowledge necessary for understanding the opportunities afforded by such an approach. As simulators become cheaper and more sophisticated it becomes increasingly likely that such a simulated hazard perception test may soon become possible and practical as part of the UK driver licensing procedure.

References

Adrian, J. V., Postal, V., Moessinger, M., Rascle, N., & Charles, A. (2011). Personality traits and executive functions related to on-road driving performance among older drivers. *Accident Analysis and Prevention, 43*(5), 1652-1659.

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychological Review, 16* (1), 17-42.

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13*(6), 909-922.

Ayaz, H., Bunce, S., Willems, B., Hah, S., Shewokis, P A., & Izzetoglu, K. (2010). Working Memory Monitoring of Air Traffic Controllers Using Functional Near Infrared Spectroscopy. *Seventeenth Annual Cognitive Neuroscience Society Meeting*, April 17-20, Montreal, Canada. URL: http://hdl.handle.net/1860/3961

Ayaz, H., Shewokis, P. A., Izzetoglu, K., Izzetoglu, M., & Onaral, B. (2011). fnirSoft: Comprehensive Signal Processing & Analysis Platform for Optical Brain Imaging. *Society of Neuroscience,* November 12-16, Washington DC, USA.

Axelrod, B. N., Goldman, R. S., Heaton, R. K., Curtiss, G., Thompson, L. L., Chelune, G. J., & Kay, G. G. (1996). Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *Journal of Clinical and Experimental Neuropsychology, 18*(3), 338-342.

Backs, R.W. & Seljos, K.A. (1994). Metabolic and cardiorespiratory measures of mental effort: The effect of level of difficulty in a working memory task. *International Journal of Psychophysiology, 16*, 57-68.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10*, 659-669.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baldo, J. V., Shimamura, A. P., Delis, D. C., Kramer, J., & Kaplan, E. (2001). Verbal and design fluency in patients with frontal lobe lesions. *Journal of the International Neuropsychological Society, 7,* 586–596.

Bandettini, P.A., Wong, E.C., Hinks, R.S., Tikofsky, R.S., & Hyde, J.S. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance In Medicine, 25*, 390-397.

Barcelo, F., & Knight, R. T. (2002). Both random and preservative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia, 40*, 349-356.

Baso-Moro, S., Cutini, S., Ursini, M. L., Ferrari, M., & Quaresima, V. (2013). Prefrontal cortex activation during story encoding/retrieval: a multi-channel functional near-infrared spectroscopy study. *Frontiers in Human Neuroscience, 7*, 925.

Beblo, T., Macek, C., Brinkers, I., Hartje, W., & Klaver, P. (2004). A new approach in clinical neuropsychology to the assessment of spatial working memory: the block suppression test. *Journal of Clinical and Experimental Neuropsychology, 26*(1), 105–114.

Bechara, A., Tranel, D., & Damasio, H. (2000) Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain, 123*(11), 2189-2202.

Bechara, A., & Van Der Linden, M. (2005). Decision-making and impulse control after frontal lobe injuries. *Current Opinion in Biology, 18*(6), 734-739.

Beeli, G., Koeneke, S., & Jancke, L. (2008). Brain stimulation modulates driving behaviour. *Behavioural and Brain Functions,* 4:34, 1-7.

Begg, D.J., & Langley, J.D. (2001). Changes in risky driving behaviour among young adults. *Journal of Safety Research, 32*, 491–499.

Benda, H.V., & Hoyos, C.G. (1983). Estimating hazards in traffic situations. *Accident Analysis & Prevention, 15*, 1-9.

Benton, A. L., & Hamsher, K. S. (1976). Multilingual Aphasia Examination: Manual of instruction. Iowa City, IA: University of Iowa.

Benton, A. L., Hamsher, K., & Sivan, A.B. (1994). Multilingual Aphasia Examination: Manual of Instruction (3rd ed.). Iowa City, IA: AJA Associates.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition, 38*, 317-338.

Bo J., Jennett S., & Seidler R. D. (2011). Working memory capacity correlates with implicit serial reaction time task performance. *Experimental Brain Research, 214*, 73–81.

Boecker, M., Buecheler, M. M., Schroeter, M. L., & Gauggel, S. (2007). Prefrontal brain activation during stop-signal response inhibition: an event-related functional near infrared spectroscopy study. *Behavioural Brain Research, 176*, 259.

Bor, D., Duncan, J., Lee, A. C. H., Parr, A., & Owen, A. M. (2006). Frontal lobe involvement in spatial span: Converging studies of normal and impaired function. *Neuropsychologia, 44*, 229-237.

Bornstein, R.A. (1986). Contribution of various neuropsychological measures to detection of frontal lobe impairment. *International Journal of Clinical Neuropsychology, 8*, 18–22.

Boucsein,W. (1992). Electrodermal Activity. New York, NY: Plenum Press.

Bouscein, W. & Backs, R.W. (2008). The Psychophysiology of Emotion, Arousal, and Personality: Methods and Models. In Duffy, V.G. (Eds.), *Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering*, (pp. 35-1 – 35-18). Boca Raton: CRC Press.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). Ananalysis of sequences of restricted associative responses. *Journal of General Psychology, 30*,149-165.

Bragg, B. W. E., & Finn, P. (1982). Younger driver risk taking: Technical report of experimental study (Rep. No. HS-806-375). Washington, DC: US Department of Transportation, NHTSA.

Braff, D., Heaton, R. K., Kuck, J., Cullum, M., Moranville, J., Grant, I., & Zisook, S. (1991). The generalised pattern of neuropsychological deficits in outpatients with chronic schizophrenia with heterogeneous Wisconsin Card Sorting Test results. *Archives of General Psychology, 48*, 891-898.

Brazy, J.E., Lewis, D.V., Mitnick, M.H., Jöbsis, & van der Vliet, F.F. (1985). Noninvasive monitoring of cerebral oxygenation in preterm infants: preliminary observations. *Pediatrics, 75*, 217–225.

Brookhuis, K.A. & de Waard, D. (2011). Measuring physiology in simulators. In Fisher, D.L, Caird, J.K., Rizzo, M., Lee, J.D. (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology: An Overview*, (pp. 17-1 - 17-10). Boca Raton: CRC Press.

Bunce, S., Izzetoglu, K., Ayaz, H., Shewokis, P A., Izzetoglu, M., Pourrezaei, K., & Onaral, B. (2011). Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload. In Dylan Schmorrow & Cali Fidopiastis (Eds.), *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems* (Vol. 6780, pp. 13-22). Berlin / Heidelberg: Springer.

Butler, R.W., Rorsman, I., Hill, J.M., & Tuma, R. (1993). The effects of frontal brain impairment on fluency: Simple and complex paradigms. *Neuropsychology, 7*, 519–529.

Calhoun V. D., Pekar, J. J., McGinty, V. B., Adali, T., Watson, T.D., & Pearlson, G. D. (2002). Different activation dynamics in multiple neural systems during simulated driving. *Human Brain Mapping, 16*(3), 158–167.

Calhoun, V. D., & Pearlson, G. D. (2012). A selective review of simulated driving studies: Combining naturalistic and hybrid paradigms, analysis approaches, and future directions. *Neuroimage, 59*(1), 25-35.

Carlesimo GA, Caltagirone C, Gainotti G & the Group for the standardization of the Mental Detrioration Battery. (1996). The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. *European Neurology, 36*, 378–384

Chance, B., Leigh, J.S., Miyake, H., Smith, D.S., Nioka, S., Greenfeld, R., …Young, M. (1988). Comparison of time-unresolved measurements of deoxyhemoglobin in brain. *Proceedings of the National Academy of Sciences U. S. A., 85*, 4971–4975.

Chapman, P., & Crundall, D. (2011). A comparison of learner drivers' responses to hazards presented as videos or animations. Report to the Driving Standards Agency.

Chapman, P.R., & Underwood, G. (1998). Visual search of driving situations: danger and experience. *Perception, 27*, 951–964.

Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F, 5*, 157–167.

Chelune, G. J., & Baer, R. A. (1986) Developmental norms for the Wisconsin Card Sorting test. *Journal of Clinical and Experimental Neuropsychology, 8*, 219–228.

Chertkow, H. & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type: What do various measures measure? *Brain, 113*, 397-417.

Chey J., Lee J., Kim Y. S., Kwon S. M., & Shin Y. M. (2002). Spatial working memory span, delayed response and executive function in schizophrenia. *Psychiatry Research, 110*, 259–271.

Chowdhury, K. & Kilbey, P. (2010). General overview and trends in reported road casualties. *In Reported Road Casualties Great Britain: Department for Transport*. London: The Stationery Office.

Clarke, D. D., Ward, P. J., & Jones, J. (1998). Overtaking road-accidents: Differences in manoeuvre as a function of driver age. *Accident Analysis & Prevention, 30*(4), 455-467.

Clarke, D.D., Ward, P., & Truman, W. (2002). In-depth accident causation study of young drivers. Report No. TRL542. Crowthorne, Berkshire: TRL Limited.

Colier, W., Quaresima, V., Barattelli, G., Cavallari, P., van der Sluijs, & M., Ferrari. (1997). Detailed evidence of cerebral hemoglobin oxygenation in response to motor cortical activation revealed by a continuous wave spectrometer with 10 Hz temporal resolution. *Proceedings of SPIE, 2979*, 390–396.

Collete, F., & van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews, 26(2)*, 105-125.

Cope, M. (1991). The application of near infrared spectroscopy to non-invasive monitoring of cerebral oxygenation in the newborn infant. PhD Thesis, University College London. URL: http://www.ucl.ac.uk/medphys/research/borl/pub/pub/theses/docs/mcope.pdf

Cope, M., & Delpy, D. T. (1988). System for long term measurement of cerebral blood and tissue oxygenation on newborn infants by near infrared transillumination *Medical & Biological Engineering and Computing, 26*(3), 289-94.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. PhD Thesis, McGill University. URL: http://digitool.library.mcgill.ca/webclient/StreamGate?folder_id=0&dvs=1467476312164~743

Crockett, D., Bilsker, D., Hurwitz, T., & Kozak, J. (1986). Clinical utility of three measures of frontal lobe dysfunction in neuropsychiatric samples. *International Journal of Neuroscience, 30*, 241–248.

Crundall, D. (2009). The Deceleration Detection Flicker Test: a measure of experience? *Ergonomics, 52* (6), 674-684.

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*, 2117–2124.

Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention, 45*, 600-609.

Crundall, D., Underwood, G., & Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology, 16*, 459–475.

Cui X., Bray S., Bryant D., Glover G. H., & Reiss A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage, 54*, 2808–2821.

S. Cutini, F. Scarpa, P. Scatturin, P. Jolicoeur, P. Pluchino, M. Zorzi & R. Dell'acqua. (2011). A hemodynamic correlate of lateralized visual short-term memories. *Neuropsychologia, 49*, 1611-1621.

Cutini, S., Scatturin, P., Menon, E., Bisiacchi, P. S., Gamberini, L., Zorzi, M., & Dell'Acqua, R. (2008). Selective activation of the superior frontal gyrus in task-switching: An event-related fNIRS study. *NeuroImage, 42*, 945.

Cummings, J. (1995). Neuropsychiatry. In: Simpson GM (Eds.). *Manual of Psychiatric Disorders* (pp. 109-136.). New York, NY: Impact Communications Inc.

Dahl, R. E. (2001). Affect regulation, brain development, and behavioural/emotional health in adolescence. *CNS Spectrums, 6*(1), 1-12.

Daigneault, G., Joly, P., & Frigon, J.Y. (2002). Previous convictions or accidents and the risk of subsequent accidents in older drivers. *Accident Analysis and Prevention, 34*, 257−261.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science, 20*, 1102-1105.

De Blasi, R.A., Fantini, S., Franceschini, M.A., Ferrari, M., & Gratton, E. (1995). Cerebral and muscle oxygen saturation measurement by frequency-domain near-infra-red spectrometer. *Medical & Biological Engineering & Computing, 33*, 228–230.

Deery, H, & Love, A. (1996). The effect of a moderate dose of alcohol on the traffic hazard perception profile of young drink drivers. *Addiction, 91*, 815-827.

Delpy, D.T., Cope, M., van der Zee, P., Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology, 33*, 1433–1442.

De Rivecourt, M., Kuperus, M.N., Post, W.J., & Mulder, L.J.M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics, 51*, 1295–1319.

de Waard, D., & Brookhuis, K.A. (1991). Assessing driver status: a demonstration experiment on the road. *Accident Analysis and Prevention, 23*, 297–307.

Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of Optical Pathlength through tissue from Direct Time of Flight Measurement. *Physics in Medicine and Biology,* 33, 1433-42.

Devinsky, O., Morrell. M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain, 118*(1), 279-306.

Dijksterhuis, C., Brookhuis, K. A., & De Waard, D. (2011). Effects of steering demand on lane keeping behaviour, self-reports, and physiology. A simulator study. *Accident Analysis and Prevention, 43*(3), 1074-1081.

Donald, M. (2001). A mind so rare. New York, NY: Norton.

Drummond, A.E. (2000). Paradigm lost! Paradise gained? An Australian's perspective on novice driver safety. *Proceedings of the Novice Drivers Conference*, 1-2 June, Bristol. Department of Transport, United Kingdom. URL: https://trid.trb.org/view.aspx?id=687213

Duncan, A., Meek, J.H., Clemence, M., Elwell, C.E., Tyszczuk, L., Cope, M., & Delpy, D.T. (1995). Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy. *Physics in Medicine and Biology, 40*, 295–304.

Duncan, J., & Owen, A.M. (2000**).** Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neuroscience, 23*(10), 475-83.

Durston, S., Davidson, M.C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development. *Developmental Science, 1*, 18–20.

Drane, D. L., Lee. G. P., Cech, M. A., Huthwaite, J. S., Ojemann, G. A., … Ojemann, J. G. (2006). Structured cueing on a semantic fluency task differentiates patients with temporal versus frontal lobe seizure onset. *Epilepsy and Behaviour, 9*(2), 339-344.

Elwell, C. E., & Cooper, C. E. (2011). Making light work: illuminating the future of biomedical optics. *Philosophical Transactions Of the Royal Society A, 369*, 4358-4379.

Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropschologia 31*(1), 17-28.

Endsley, M. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors*, 37, 65-84.

Ettlinger, G., Teuber, H., & Milner, B. (1975). The seventeenth International Symposium of Neuropsychology. *Neuropsychologia, 13*, 125–133.

European Conference of Ministers of Transport. (2006). Improving Transport Accessibility For All: Guide to Good Practice. Organisation for Economic Co-operation and Development. DOI: 10.1787/9789282101407-en

Fecteau, S., Knoch, D., Fregni, F., Sultani, N., Boggio, P., & Pascual-Leone, A. (2007). Diminishing risk-taking behaviour by modulating activity in the prefrontal cortex: a direct current stimulation study. *Journal of Neuroscience*, *27*(46), 12500-12505.

Ferrari, M., Giannini, I., Sideri, G., & Zanette, E. (1985). Continuous non invasive monitoring of human brain by near infrared spectroscopy. *Advances in Experimental Medicine and Biology, 191*, 873–882.

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage, 63*(2), 921-935.

Field, D. T., Groeger, J. A., & Hammond, S. M. (1997). Computerised assessment of drivers' skills. In G B Grayson (Ed.) *Behavioural Research in Road Safety VII* (pp.49-59). Crowthorne, UK: TRL Limited.

Fillmore MT, Blackburn JS, & Harrison ELR. (2008). Acute disinhibiting effects of alcohol as a factor in risky driving behaviour. *Drug and Alcohol Dependence, 95*, 97–106.

Fischer M. H. (2001). Probing spatial working memory with the Corsi Blocks Task. *Brain and Cognition, 45*, 143–154.

Fitzgerald, E.S., & Harrison,W.A. (1999). Hazard perception and learner drivers: A theoretical discussion and an in-depth survey of driving instructors. Report No. 161. Melbourne: Monash University Accident Research Centre. URL:http://www.monash.edu/__data/assets/pdf_file/0007/217375/muarc161.pdf

fNIR Imager & COBI Studio Manual. Cognitive Optical Brain Imaging Studio. fNIR Devices LLC. Ptomac, MD. URL: https://www.biopac.com/wp-content/uploads/fnir-cobi-manual.pdf

Forsyth, E. (1992). Cohort study of learner and novice drivers. Part 1. Learning to drive and performance in the driving test. Report No. RR338. Crowthorne, UK: Transport Research Laboratory. URL: http://www.trl.co.uk/reports-publications/trl-reports/report/?reportid=5938

Fox, P.T., & Raichle, M.E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences. U. S. A.*, *83*, 1140 – 1144.

Franceschini, M.A., Toronov, V., Filiaci, M., Gratton, E., & Fantini, S. (2000). On-line optical imaging of the human brain with 160-ms temporal resolution. *Optics Express, 6*, 49–57.

Fuster, J. M. (1997). The Prefrontal Cortex-Anatomy Physiology, and Neuropsychology of the Frontal Lobe (3rd ed.). Philadelphia, PA: Lippincott-Raven.

Fuster, J. M., Bodner, M., & Kroger, J. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature, 405*, 347-351.

Garay, L., Fisher, D.l., & Hancock, K.L. (2004). Effects of driving experience and lighting condition on driving performance. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 48*, 2290–2294.

Garay-Vega, L., Fisher, D.L., & Pollatsek, A. (2007). Hazard anticipation of novice and experienced drivers: empirical evaluation on a driving simulator in daytime and nighttime conditions. *Transportation Research Record: Journal of the Transportation Research Board, 2009*, 1–7.

Gefen, D., Ayaz, H., & Onaral, B. (2014). Applying Functional Near Infrared (fNIR) Spectroscopy to Enhance MIS Research. *AIS Transactions on Human-Computer Interaction 6*(3), 55-73.

Gevins, A., Leong, H., Smith, M. E., Le, J., & Du, R. (1995). Mapping cognitive brain function with modern high resolution electroencephalography. *Trends in Neuroscience, 18*(10). 429-436.

Graham, K.J., & Kinney, G.F. (1980) A practical safety analysis system for hazards control. *Journal of Safety Research, 12*, 13-20.

Grant, D. A., & Berg, E. A. (1948). A behavioural analysis of degree reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology 38*, 404-441.

Grafman, J., Jonas, B., & Salazar, A. (1990). Wisconsin Card Sorting Test performance based on location and size of neuroanatomical lesion in Vietnam veterans with penetrating brain injuries. *Perceptual and Motor Skills, 71*, 1120-1122.

Gratton, G., Corballis, P.M., Cho, E., Fabiani, M., & Hood, D.C. (1995). Shades of gray matter: noninvasive optical images of human brain responses during visual stimulation. *Psychophysiology, 32*, 505–509.

Graydon, F. X., Young, R. A., Genik, R. J, II., Posse, S., Hsieh L, & Green, C. C. (2004). Visual event detection during simulated driving: Identifying the neural correlates with functional neuroimaging. *Transportation Research Part F: Traffic Psychology and Behaviour, 7*, 271.

Grayson, G. B., & Maycock, G. (1997). What do drivers learn through experience - and can it be taught? Paper presented to NHTSA Symposium on Risk Taking Behavior and Traffic Safety. Cape Cod, October 1997.

Grayson, G. B., & Sexton, B. F. (2002). The development of hazard perception testing. Prepared for Road Safety Division, Department of Transport. TRL Report 558. Crawthorne, England: Transport Research Laboratory. URL: http://static.a2om.com.s3.amazonaws.com/driveiq/more-research/hpt%20research/TRLeshptreseach%20project.pdf

Gruenewald, P. & Lockhead, G.R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 225-240.

Haines, J., Hoffman, W., Hanada. J., Brown, P., Labs, S., Loberg, D., & Walter, C. T. (1994). Further evaluation of prefrontal lobe function in schizophrenic patients during Wisconsin Card Sorting Test (letter). *American Journal of Psychiatry 151*(12), 1842-1843.

Hamilton, J. W. (1860). Editorial and Miscellaneous. The Man Through Whose Head an Iron Rod Passed Is Still Living. *Ohio Medical and Surgical Journal 13*, 174.

Harlow, J. M. (1848). Passage of an iron rod through the head. *Boston Medical & Surgical Journal, 39*(20), 389-393.

Harlow, J. M. (1868). Recovery from the passage of an iron bar through the head. *Publications of the Massachusetts Medical Society, 2*, 327-347.

Harada H., Nashihara H., Morozumi K., Ota H., & Hatakeyama E. (2007). A comparison of cerebral activity in the prefrontal region between young adults and the elderly while driving. *Journal of Physiological Anthropology 26*, 409–414.

Harré, N., Field, J., & Kirkwood, B. (1996). Gender differences and areas of common concern in the driving behaviors and attitudes of adolescents. *Journal of Safety Research, 27*(3), 163-173.

Harrington, D. M., & McBride, R. S. (1970). Traffic violations by type, age, sex and marital status. *Accident Analysis & Prevention, 2*(1), 67-79.

Hart, S., & Staveland, L. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland Press.

Hatakenaka, M., Miyai, I., Mihara, M., Sakoda, S. & Kubota, K. (2007). Frontal regions involved in learning of motor skill-A functional NIRS study. *Neuroimage, 34*(1), 109-16.

Haxby J. V., Lundgren S. L., & Morley G. K. (1983). Short-term retention of verbal, visual shape and visuospatial location information in normal and amnesic subjects. *Neuropsychologia 21*, 25–33.

Hedden, T., & Yoon, C. (2006). Individual Differences in Executive Processing Predict Susceptibility to Interference in Verbal Working Memory. *Neuropsychology, 20 (5)*, 511-528.

Healey, J.A. & Picard, R.W. (2005). Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems, 6* (2), 156–166.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). Wisconsin Card Sorting Test (WCST). Manual revised and expanded. Odessa, FL: Psychological Assessment Resources Inc.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia 42*, 1212–1222.

Herrmann, M. J., Plichta, M. M., Ehlis, A.-C., & Fallgatter, A. J. (2005). Optical topography during a Go-NoGo task assessed with multi-channel near-infrared spectroscopy. *Behavioural Brain Research, 160*, 135.

Herrmann, M. J., Walter, A., Ehlis, A.-C. & Fallgatter, A.J. (2006). Cerebral oxygenation changes in the prefrontal cortex: Effects of age and gender. *Neurobiology of Aging 27*, 888.

Hinson, J. M., Jameson, T. L., & Whitney, P. 2003. Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(2): 298-306.

Hiraoka, M., Firbank, M., Essenpreis, M., Cope, M., Arridge, S. R., van der Zee, P., & Delpy, D. T. (1993) A Monte Carlo investigation of optical pathlength in homogeneous tissue and its application to near-infrared spectroscopy. *Physics in Medicine and Biology*, *38*, 1859−1876.

Haworth, N, Symmons, M, & Kowadlo, N. (2000). Hazard Perception by Inexperienced Motorcyclists. Report No. 179 Melbourne: Monash University Accident Research Centre. URL: https://www.monash.edu/__data/assets/pdf_file/0005/217454/muarc179.pdf

Holland, C., Geraghty, J., & Shah, K. (2010). Differential moderating effect, of locus of control on effect of driving experience in young male and female drivers. *Personality and individual differences, 48*(7), 821-826.

Holper, L., Biallas, M., & Wolf, M. (2009). Task complexity relates to activation of cortical motor areas during uni and bimanual performance: a functional NIRS study. *NeuroImage, 46*, 1105-1113.

Horikawa E., Okamura N., Tashiro M., Sakurada Y., Maruyama M., Arai H., … Itoh, M. (2005). The neural correlates of driving performance identified using positron emission tomography. *Brain and Cognition, 58*, 166–171.

Hoshi, Y., & Tamura, M. (1993). Detection of dynamic changes in cerebral oxygenation coupled to neuronal function mental work in man. *Neuroscience Letters,* 150, 5–8.

Hoshi, Y., & Tamura, M. (1997). Near-infrared optical detection of sequential brain activation in the prefrontal cortex during mental tasks. *NeuroImage, 5*, 292-297.

Isler, R. B., Starkey, N. J., Drew, M., & Sheppard, P. (2008). A double-blind, randomized controlled study of the effectiveness of higher level driving skills training to improve frontal lobe (executive) function related driving performance in young drivers. Commissioned report for the AA Driver Education Foundation. URL: http://www.waikato.ac.nz/__data/assets/pdf_file/0017/142208/Frontal-Lobe-June2008.pdf

Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., & Pourrezaei, K. (2005). Functional Near-Infrared Neuroimaging. *IEEE Transactions On Neural Systems And Rehabilitation Engineering, 13*(2), 153-159.

Jäncke, L., Brunner, B., & Esslen, M. (2008). Brain activation during fast driving in a driving simulator: The role of the lateral prefrontal cortex. *Neuro Report, 19*, 1127–1130.

Jöbsis, F.F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science, 198*, 1264–1267.

Jongen, E. M. M., Brijs, K., Komlos, M., Brijs, T., & Wets, G. (2011). Inhibitory control and reward predict risky driving in young novice drivers – a simulator study. Proceedings of the sixth international driving symposium on human factors in driver assessment, training and vehicle design, 27–30 June, Lake Tahoe, California, USA. URL: http://drivingassessment.uiowa.edu/sites/default/files/DA2011/Papers/078_JongenBrijs.pdf

Jonowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioural Neuroscience, 103*(3), 548-560.

Just, M. A., Keller, T. A., & Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain Research, 1205*, 70–80.

Kaneko, H., Yoshikawa, T., Nomura, K., Ito, H.,Yamauchi, H., Ogura M., & Honjo, S. (2011). Hemodynamic changes in the prefrontal cortex during digit span task: A near-infrared spectroscopy study. *Neuropsychobiology, 63*, 59.

Kato, T., Kamei, A., Takashima, S., & Ozaki, T. (1993). Human visual cortical function during photic stimulation monitoring by means of near-infrared spectroscopy. *Journal of Cerebral Blood Flow and Metabolism, 13*, 516–520.

Kawasaki, Y., Maeda, Y., Suzuki, M., Urata, K., Higashima, M., Kiba, K., …Hisada, K. (1993). SPECT analysis of regional cerebral blood flow changes in patients with schizophrenia during the Wisconsin Card Sorting Test. *Schizophrenia Research 10*, 109–116.

Kawashima, R., Nouchi, R., Matsumoto, T., & Tanimoto, Y. (2014). Riding a motorcycle affects cognitive functions in healthy adults. *International Journal of Automotive Engineering, 5*, 73-76,

Keating D. (2004). Cognitive and brain development. *In: Lerner R, Steinberg L, (eds). Handbook of adolescent psychology*. (Vol. 2, pp. 45–84.) New York: Wiley.

Kessels, R. P. C., Van Zandvoort, M. J. E., Postma, A., & De Haan, E. H. F. (2000). The Corsi block-tapping task: Standardisation and normative data. *Applied Neuropsychology, 7*, 252–258.

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (Vol. 5, pp. 47–92). Greenwich, CT: JAI Press.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (1993). Decision making in action: Models and methods. Norwood, NJ: Ablex Publishing Corporation.

Koch, S.P., Habermehl, C., Mehnert, J., Schmitz, C.H., Holtze, S., Villringer, A., … Obrig, H. (2010). High-resolution optical functional mapping of the human somatosensory cortex. *Frontiers in Neuroenergetics, 2*, 12.

Kocsis, L., Herman, P., and Eke, A. (2006). The Modified Beer-Lambert Law Revisited. *Physics in Medicine and Biology,* 51, 91-98.

Konstantopoulos, P., Crundall, D., & Chapman, P. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore visual search in day, night and rain driving. *Accident Analysis and Prevention, 42*, 827–834.

Kojima, T., Tsunashima, H., Shiozawa, T., Takada, H., and Sakai, T. (2005). Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS). *Optical and Quantum Electronics, 37*, 13–15.

Kovelman, I., Shalinsky, M.H., Berens, M.S., & Petitto, L.A. (2008). Shining light on the brain's "bilingual signature": A functional Near Infrared Spectroscopy investigation of semantic processing. *NeuroImage, 39*, 1457 - 1471.

Krawczyk, D. C. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neuroscience & Biobehavioral Reviews, 26*, 631-664.

Lajunen, T., and Parker, D. (2001). Are aggressive people aggressive drivers? A study of the relationship between self-reported general aggressiveness, driver anger and aggressive driving. *Accident Analysis & Prevention, 33*(2), 243-255.

Langford, J., Koppel, S., McCarthy, D., & Srinivasan, (2008). In defense of the 'low-mileage bias'. *Accident Analysis and Prevention, 40*(6), 1996-1999.

Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage, 140*, 1044-1055.

Lee, J. D. (2007). Technology and teen drivers. *Journal of Safety Research, 38*(2), 203-213.

Lee, S. E., Klauer, S. G., Olsen, E. C. B., Simons-Morton, B. G., Dingus, T. A., Ramsey, D. J., Ouimet, M. C. (2008). Detection of road hazards by novice teen and experienced adult drivers. *Transportation Research Record*, *2078*, 26–32.

Lee and, M. F., Lee, D. N. (1994). Where we look when we steer. *Nature, 369*, 742-744.

Lerner, N., & Rabinovich, B. (1997). Risk Perception in young male drivers: what makes them different? *Proceedings of the Human Factors & Ergonomics Society Annual Meeting, 41*(2), 993-997.

Levin, H. S., Culhane, K. A., & Hartmann, J. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 7*, 377–395.

Levy, D. T. (1990). Youth and traffic safety: the effects of driving age, experience and education. *Accident Analysis and Prevention, 22*(4), 327-334.

Lewis-Evans, B., & Lukkien, C. (2007). Crash profile of New Zealand novice drivers. Australasian Road Safety Research Policing Education Conference. Melbourne, Victoria, Australia. URL: http://acrs.org.au/files/arsrpe/RS07039.pdf

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). Neuropsychological assessment (4th ed.). New York, NY: Oxford University Press.

Lie, C. H., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *Neuroimage, 30*, 1038–1049.

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Hazard perception abilities of experienced and novice motorcyclists: An interactive simulator experiment. *Transportation Research Part F: Psychology and Behaviour, 12*, 325–334.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*, 150-157.

Lonczak, H. S., Neighbors , C., & Donovan, D. M. (2006). Predicting risky and angry driving as a function of gender. *Accident Analysis and Prevention, 39*(3), 536-545.

Luck, S. J. (2005). An Introduction to the Event-Related Potential Technique. Cambridge, MA: The MIT Press.

Luna, B., Thulborn, K. R., Munoz, D. P., Merriman, E. P., Garver, K. E., Minshew, N. J., … Sweeney, J. A. (2001). Maturation of widely distributed brain function subserves cognitive development. *NeuroImage, 13*, 786 –793.

Luria, A. R. (1966). Higher cortical functions in man. New York, NY: Basic Books.

Luria, A. R. (1973). The Working Brain. London: Penguin Books.

Lezak, M. D. (2004). Neuropsychological Assessment (2nd ed.). New York, NY: Oxford University Press.

Maki, A., Yamashita, Y., Ito, Y., Watanabe, E., Mayanagi, Y., & Koizumi, H. (1995). Spatial and temporal analysis of human motor activity using noninvasive NIR topography. *Medical Physics, 22*, 1997–2005.

Malonek, D. & Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy implications for functional brain mapping. *Science, 272*, 551 – 554.

Matthews, M. L., & Moran, A. R. (1986). Age differences in male drivers perception of accident risk: the role of perceived driving abilities. *Accident Analysis and Prevention, 18*(4), 299-313.

Maycock, G. (2002). Novice driver accidents and the driving test. TRL Report, TRL527. Crowthorne, UK: Transport Research Laboratory. URL: http://www.trl.co.uk/reports-publications/report/?reportid=2681

Maycock G, Lockwood, C. R., & Lester, J. F. (1991). The accident liability of car drivers. TRRL Research Report RR315. Crowthorne, UK: Transport and Road Research Laboratory.

Mayhew, D. R., Simpson, H. M., Williams, A. F., & Ferguson, S. A. (1998). Effectiveness and role of driver education and training in a graduated licensing system. *Journal of Health Policy, 19*(1), 51-67.

Mayhew, D. R., Simpson, H. M., & Pak, A. (2003). Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis & Prevention, 35*(5), 683-691.

Meek, J.H., Elwell, C.E., Khan, M.J., Romaya, J., Wyatt, J.S., Delpy, D.T., & Zeki, S. (1995). Regional changes in cerebral haemodynamics as a result of a visual stimulus measured by near infrared spectroscopy. *Proceedings of the Royal Society B: Biological Sciences, 261*, 351–356.

Mehler, B., Reimer, B. & Coughlin, J.F. (2012). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand from a Working Memory Task: An On-road Study across Three Age Groups. Human Factors, *54*(3), 396-412

Mentzel, H.J., Gaser, C., Volz, H.P., Rzanny, R., Hager, F., Sauer,H., & Kaiser, W.A. (1998). Cognitive stimulation with the Wisconsin Card Sorting Test: Functional MR imaging at 1.5 T. *Radiology, 207*, 399-404.

McCartt, A. T., Mayhew, D. R., Braitman, K. A., Ferguson, S. A., & Simpson, H. M. (2009). Effects of age and experience on young driver crashes: Review of recent literature. *Traffic Injury Prevention, 10*(3), 209-219.

Mcdonald W. A. (1994). Young Drivers Research Program: A Review on Information on Young Driver Crashes. Federal Office of Road Safety, Australia.URL: http://www.monash.edu/__data/assets/pdf_file/0020/217271/atsb128.pdf

McGowan A, & Banbury S. (2004). Evaluating interruption-based techniques using embedded measures of driver anticipation. In: Banbury, S. & Tremblay, S., A *cognitive approach to situation awareness*. Aldershot, UK: Ashgate.

McKenna, F. (2000). What is going wrong in the early part of the driving career? Proceedings of the Novice Drivers Conference, 1-2 June, Bristol. London: Department of Transport.

McKenna, F.P., Crick, J.L., (1994). Hazard perception in drivers: a methodology for testing and training. TRL Report 313. Crowthorne, UK: Transport Research Laboratory. URL: http://www.trl.co.uk/reports-publications/report/?reportid=3121

McKenna, F.P., & Horswill, M.S. (1999). The development, validation and application of video based technique for measuring an everyday risk-taking behaviour: drivers' speed choice. *Journal of Applied Psychology 84*, 977–985.

McKinstry, R. C., Mathur, A., Miller, J. H., Ozcan, A., Snyder, A. Z., Schefft, G. L., … Neil, J. J. (2002). Radial organisation of developing preterm human cerebral cortex revealed by non-invasive water diffusion anisotropy MRI. *Cerebral Cortex, 12*(12), 1237-1243.

McKnight and McKnight. (2003). Young novice drivers: careless or clueless? *Accident Analysis and Prevention, 35*, 921-923.

Milech, D., Glencross, D., & Hartley, L. (1989). Skills acquisition by young drivers: Perceiving, interpreting and responding to the driver environment. Report MR 4. Canberra: Australia. Federal Office of Road Safety. URL: https://infrastructure.gov.au/roads/safety/publications/1989/pdf/mr_04.pdf

Miller, B. L., & Cummings, J. L. (2007). The human frontal lobes: functions and disorders. New York, NY: The Guilford Press.

Mills, K.C. (2005). Disciplined Attention: How to Improve Your Visual Attention When You Drive. Chapel Hill, NC: Profile Press.

Mills, K.L, Rolls, G.W.P., Hall, R.D., & McDonald, M. (1995). The effects of hazard perception training on the development of novice driver skills. Final Report to the Department of Transport.

Milner, B. (1963). Effects of different brain lesions on card-sorting: The role of the frontal lobes. *Archives of Neurology 9*, 100-110.

Milner, B. (1964). Some effects of frontal lobectomy in man. In: Warren, J.M. and Akert,K. (Eds.), *The frontal granular cortex and behaviour* (pp. 313–335). New York: McGraw Hill.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*, 272–277.

Miyai, I., Tanabe, H.C., Sase, I., Eda, H., Oda, I., Konishi, I., Kubota, K. (2001). Cortical mapping of gait in humans: a near-infrared spectroscopic topography study. *NeuroImage, 14*, 1186–1192.

Monchi, O., Petrides, M. Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, *21*(19): 7733-7741.

Moscovitch, M., & Winocur, G. (1995). Frontal lobes, memory, and aging. *Annals of the New York Academy of Sciences,769*, 119–150.

Mourant, R.R., & Rockwell, T.H. (1972). Strategies of Visual Search by Novice and Experienced Drivers. *Human Factors*, 325-335.

Nagahama, Y., Fukuyama, H., Yamauchi, H., Matsuzaki, S., Konishi, J., Shibasaki, H., & Kimura, J. (1996). Cerebral activation during performance of a card sorting test. *Brain 119*, 1667-1675.

Naatanen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention, 6*(3-4): 243-261.

Nakahachi, T., Ishii, R., Iwase, M., Canuet, L., Takahashi, H., & Kurimoto, R. (2010). Frontal cortex activation associated with speeded processing of visuospatial working memory revealed by multichannel near-infrared spectroscopy during Advanced Trail Making Test performance. *Behavioural Brain Research, 215*, 21-27.

New Zealand Ministry of Transport. (2010). Yearly Report: Motor Vehicle Crashes In New Zealand 2009. Wellington, New Zealand. URL: http://www.transport.govt.nz/assets/Import/Documents/Motor-Vehicle-Crashes-in-New-Zealand-2009.pdf

NHTSA. (2009). Traffic Safety Facts 2009. Washington, DC: US Department of Transportation, NHTSA. URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811402

Niedermeyer E. & da Silva F.L. (2004). Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Philadelphia, PA: Lippincot Williams & Wilkins

Noguchi Y, Takeuchi T, Sakai KL (2002). Localized activation in the inferior frontal cortex during syntactic processing: an event-related optical topography study. *Human Brain Mapping, 17, 89–99*

O'Brien, F., & Gormley, M. (2013). The contribution of inhibitory deficits to dangerous driving among young people, *Accident Analysis and Prevention, 51*, 238 – 242.

Obrig, H., Villringer, A. (2003). Beyond the visible-imaging the human brain with light. *Journal of Cerebral Blood Flow & Metabolism, 23*, 1–18.

OECD/ECMT Transport Research Centre. (2006). Young Drivers: The Road to Safety. Paris: OECD Publishing. URL: http://www.oecd.org/itf/37556934.pdf

Okada, E., Firbank, M., Schweiger,M., Arridge, S.R., Cope, M., & Delpy, D.T., (1997). Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. *Applied Optics, 36*, 21–31.

Owen, A. M., Evans, A. C., & Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: A positron emission tomography study. *Cerebral Cortex, 6*, 31-38.

Pardo, J. V., Fox, P. T., & Raichle, M. E. (1994). Localisation of a human system for sustained attention by positron emission tomography. *Nature, 3*(349), 61-64.

Peres M., van de Moortele P. F., Pierard C., Lehericy S., LeBihan D., & Guezennec C. Y. (2000). fMRI of mental strategy in a simulated aviation performance task. *Aviation Space and Environmental Medicine, 71*, 1218–1231

Pollatsek, A., Narayanaan, V., Pradhan, A., &Fisher, D.L. (2006). Using eye movementsto evaluate a PC-based risk awareness and perception training program on a driving simulator. *Human Factors, 48* (3), 447–464.

Pradhan, A.K., Hammel, K.R., DeRamus, R., Pollatsek, A., Noyce, D.A., & Fisher, D.L. (2005). Using eye movements to evaluate effects of driver age on risk perception in a driving simulator. *Human Factors, 47*, 840–852.

Pradhan, A.K., Pollatsek, A., Knodler, M., & Fisher, D.L. (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway scenarios that are hard to identify as hazardous? *Ergonomics, 52*, 657–673.

Quaresima, V., Ferrari, M., van der Sluijs, M.C., Menssen, J., & Colier, W.N. (2002). Lateral frontal cortex oxygenation changes during translation and language switching revealed by non-invasive near-infrared multi-point measurements. *Brain Research Bulletin, 59*, 235–243.

Quimby, A.R., Maycock, G., Carter, I.D., Dixon, R., & Wall, J.G., (1986). Perceptual abilities of accident involved drivers. TRL Report RR27. Crowthorne, UK: Transportation and Road Research Laboratory. URL: http://www.trl.co.uk/reports-publications/trl-reports/report/?reportid=5668

Quimby, A. R., & Watts, G. R. (1981). Human factors and driving performance. Laboratory Report 1004. Crowthorne, UK: Transport and Road Research Laboratory. URL: http://www.trl.co.uk/reports-publications/trl-reports/report/?reportid=4563

Raikos, M.K. (2003). Improving speed in skilled anticipation: the case of hazard perception in driving. Unpublished Honors Thesis. University of Queensland, Brisbane.

Reed, M. P., & Green, P. A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialing task. *Ergonomics, 42*, 1015–1037.

Retzlaff, P., Butler, M., & Vanderploeg, R. D. (1970). Neuropsychological battery choice and theoretical orientations: a multivariate analysis. *Journal of Clinical Psychology, 48*, 666-672.

Rezai, K., Andreasen, N.C., Alliger, R., & Cohen, G. (1993). The neuropsychology of the prefrontal cortex. *Archives in Neurology, 50*, 636–642.

Rhodes, N., & Pivik, K. (2011). Age and gender differences in risky driving: The roles of positive affect and risk perception. *Accident Analysis & Prevention, 43*, 923-931.

Rich, J. B., Troyer, A. K., Bylsma, F. W., & Brandt, J. (1999). Longitudinal analysis of phonemic clustering and switching during word-list generation in Huntington's disease. *Neuropsychology, 13*(4), 525-531.

Richardson, E. D., & Marottoli, R. A. (2003). Visual attention and driving behaviors among community-living older persons. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 58*, 832–836.

Robert, P. H., Lafont, V., Medecin, I., Berthet, L., Thauby, S., Baudu, C., & Darcourt, G. (1998). Clustering and switching strategies in verbal fluency tasks: Comparison between schizophrenics and healthy adults. *Journal of the International Neurological Society, 4*, 539-546.

Rolls, E. T. (1994). Neurophysiology and cognitive functions of the striatum. *Revue Neurologique (Paris), 150*, 648-660.

Ross, T. P. (2003). The reliability of cluster and switch scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology, 18*(2), 153-164.

Ross, T, P., Calhoun, E., Cox, T., Wenner, C., Kono, W., & Pleasant, M. (2007). The reliability and validity of qualitative scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology, 22*, 475-488.

Rubia, K., Overmayer, S., Taylor, E., Brammer, M., Williams, S. C., Simmons, A., … Bullmore, E. T. (2000). Functional frontalisation with age: mapping neurodevelopmental trajectories with fMRI. *Neuroscience & Biobehavioral Reviews, 24*(1), 13-19.

Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience 23*, 1752.

Ruff, R. M., Light, H., Parker, S. B., & Levin, H. S. (1996). Benton Controlled Oral Word Association Test: Reliability and Updated Norms. *Archives of Clinical Neuropsychology, 11(4)*, 329-338.

Ryan GA, Legge M, & Rosman D. (1998). Age related changes in drivers' crash risk and crash type. *Accident Analysis & Prevention, 30*(3), 379–387.

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis and Prevention, 382*, 407–414.

Sakai, H., Uchimaya, Y., Tanaka, S., Sugawara, S. K., & Sadato, N. (2014). Prefrontal transcranial direct current stimulation improves fundamental vehicle control abilities. *Behavioural Brain Research, 15*, 57-62.

Salthouse, T.A., Atkinson, T.M., & Berish, D.E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566–594.

Sato, T., Ito, M., Suto, T., Kameyama, M., Suda, M., & Yamagishi, Y. (2007). Time courses of brain activation and their implications for function: a multichannel near-infrared spectroscopy study during finger tapping. *Neuroscience Research, 58*, 297.

Sbodorne, R. J. (2000). In Groth-Marnat (Ed.) *The executive functions of the brain, in neuropsychological assessment and clinical practice: A guide to test interpretation and integration* (pp. 437-456). New York, NY: Wiley.

Schreppel, T., Egetemeir, J., Schecklmann, M., Plichta, M.M., Pauli, P., & Ellgring, H. (2008). Activation of the prefrontal cortex in working memory and interference resolution processes assessed with near-infrared spectroscopy. *Neuropsychobiology, 57*, 188–193.

Senserrick, T. M. (2006). Reducing young driver road trauma: guidance and optimism for the future. *Injury Prevention, 12*, 56-60.

Shechtman, O., Classen, S., Awadzi, K., & Mann, W. (2009). Comparison of driving errors between on-the-road and simulated driving assessment: A validation study. *Traffic Injury Prevention, 10*, 379–385.

Schecklmann, M., Ehlis, A.-C., Plitcha, M. M., & Fallgatter, A. J. (2008). Functional near-infrared spectroscopy: A long-term reliable tool for measuring brain activity during verbal fluency. *NeuroImage, 43*(1), 147-155.

Schieber, F., Schlorholtz, B., & McCall, R. (2009). Visual requirements of vehicular guidance. In C. Castro (Ed.) *Human factors of visual and cognitive performance in driving* (pp. 31–50). Oxford, UK: Taylor and Francis.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia J. M., Wolf U., & Wolf, M. (2013). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage, 85*, 6–27.

Schroeter, M. L., Zysset, S., Kupka, T., Kruggel, F., & D. Cramon, Yves von. (2002). Near-infrared spectroscopy can detect brain activity during a color-word matching Stroop task in an event-related design. *Human Brain Mapping, 17*, 61.

Sowell, E. R., Thompson, P M., Holmes, C. J., Jernigan, T. L., & Toga, A. R. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience, 2*(10), 859-861.

Sowell E. R., Thompson P. M., Tessner K. D., & Toga A. W. (*2001*). Mapping continued brain growth and gray matter density reduction in dorsal frontal cortex: Inverse relationships during post adolescent brain maturation. *Journal of Neuroscience, 21,* 8819-*8829.*

Shahar, A., Poulter, D., Clarke, D., & Crundall, D. (2010). Motorcyclists' and car drivers' responses to hazards. *Transportation Research F: Psychology and Behaviour, 13,* 243–254.

Shallice, T. (1988). From neuropsychology to mental structure. New York, NY: Cambridge University Press.

Smyth, M.M., & Scholey, K.A. (1994). Characteristics of spatial memory span: Is there an analogy to the word length effect, based on movement time? *Quarterly Journal of Experimental Psychology, 47A*, 91-117.

Spiers H. J., & Maguire E. A. (2007). Neural substrates of driving behaviour. *Neuroimage 36*, 245–255.

Spreen, O., & Strauss, E. (1998). A compendium of neuropsychological tests: Administration, norms and commentary, 2nd edition. New York, NY: Oxford University Press.

Srinivasan, R. (1999). Methods to improve the spatial resolution of EEG. *International Journal of Bioelectromagnetism, 1*(1). 102-111.

Steinberg, L. (2007). Risk Taking in Adolescence New Perspectives From Brain and Behavioral Science. *Current Directions in Psychological Science 16*(2), 55-59.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28*, 78-106.

Strangman, G., Culver, J. P., Thompson, J. H., & Boas, D. A. (2002). A quantitative comparison of BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage, 17*(2), 719-731.

Stuss, D. T., Alexander, M. P., Hamer, L., Palumbo, C., Dempster, R., Binns, M., et al. (1998). The effects of focal anterior and posterior brain lesions on verbal fluency. *Journal of the International Neuropsychological Society*, *4*, *265*–78.

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95*, 3-28.

Stuss, D.T., & Knight, R.T. (2002). Principles of frontal lobe function. New York, NY: Oxford University Press.

Suzuki M., Miyai I., Ono T., Oda I., Konishi I., Kochiyama T., & Kubota, K. (2004). Prefrontal and premotor cortices are involved in adapting walking and running speed on the treadmill: an optical imaging study. *Neuroimage, 23*, 1020–1026.

Taylor, D. H. (1964). Driver's galvanic skin response and the risk of accident. *Ergonomics, 7*, 439-451.

Teixeira Ferreira, C., Verin, M., Pillon, B., Levy, R., Dubois, B., & Agid, Y. (1998). Spatio-temporal working memory and frontal lesions in man. *Cortex, 34*, 83-98.

Tien, A.Y., Schlaepfer, T.E., Orr, W., & Pearlson, G.D. (1998). SPECT brain blood flow changes with continuous ligand infusion during previously learned WCST performance. *Psychiatry Research, 82*, 47-52.

Toepper, M., Beblo, T., Thomas, C., & Driessen, M. (2008). Early detection of Alzheimer's disease: A new working memory paradigm. *International Journal of Geriatric Psychiatry, 23(3)*, 272–278.

Toepper, M., Gebhardt, H., Beblo, T., Thomas, C., Driessen, M., Bischoff, M., … Sammer, G. (2010). Functional correlates of distractor suppression during spatial working memory encoding. *Neuroscience, 165*, 1244-1253.

Toichi M, Findling RL, Kubota Y, Calabrese JR, Wiznitzer M, McNamara NK, & Yamamoto K. (2004). Hemodynamic differences in the activation of the prefrontal cortex: attention vs. higher cognitive processing. *Neuropsychologia, 42*, 698–706.

Tombaugh, N. T., Kozak, J., & Rees, L. (1999). Normative Data Stratified by Age and Education for Two Measures of Verbal Fluency: FAS and Animal Naming. *Archives of Clinical Neuropsychology, 14(2)*, 167-177.

Torricelli, A., Contini, D., Pifferi, A., Caffini, M., Re, R., Zucchelli, L., & Spinelli, L. (2015). Time domain functional NIRS imaging for human brain mapping. *NeuroImage, 85*(1), 28-50.

Troyer, A. K. (2000). Normative Data for Clustering and Switching on Verbal Fluency Tasks. *Journal of Clinical and Experimental Neuropsychology, 22(3)*, 370-378.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology, 11(1)*, 138-146.

Troyer, A. K., Moscovich, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal-and temporal-lobe lesions. *Neuropsychologia, 36*, 449–504.

Tsunashima, H., and Yanagisawa, K. (2009). Measurement of brain function of car river using functional near-infrared spectroscopy (fNIRS). *Computational Intelligence and Neuroscience*, *2009*, 1-12.

van Asselen, M., van der Lubbe, R. H. J., & Postma, A. (2006). Are space and time automatically integrated in episodic memory? *Memory, 14*, 232-240.

Uchiyama Y., Ebe K., Kozato A., Okada T., & Sadato N. (2003). The neural substrates of driving at a safe distance: a functional MRI study. *Neuroscience Letters, 352*, 199–202.

Wallace, P., Haworth, N. and Regan, M. (2005). Best methods for teaching hazard perception and responding by motorcyclists. Report No. 236. Melbourne: Monash University Accident Research Centre. URL: https://www.monash.edu/__data/assets/pdf_file/0017/217214/muarc236.pdf

Waller, P. F., Elliott, M. R., Shope, J. T., Raghunnathan, T. E., & Little, R. J. (2001). Changes in young adult offense and crash patterns over time. *Accident Analysis and Prevention, 33*(1), 117-128.

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*, 1137–1185.

Walter H, Vetter SC, Grothe J, Wunderlich AP, Hahn S, and Spitzer M. (2001). The neural correlates of driving. *Neuroreport, 12*(8),1763–1767.

Warkentin, S., & Passant, U. (1997). Functional imaging of the frontal lobes in organic dementia. *Dementia & Geriatric Cognitive Disorders, 8*, 105–109.

Watts, G.R., & Quimby, A.R. (1979). Design and validation of a driving simulator for use in perceptual studies. Laboratory Report LR907. Crowthorne, UK: Transport and Road Research Laboratory. URL: http://www.trl.co.uk/reports-publications/trl-reports/road-safety/report/?reportid=4458

Weinberger, D.R., Berman, K.F., & Zee, R.F. (1986). Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia. *Archives in General Psychiatry, 43*, 114–124.

Weinberger, D. R., Elvevag, B., & Giedd, J. N. (2005). The adolescent brain: a work in progress. The National Campaign to Prevent Teen Pregnancy. URL: https://thenationalcampaign.org/sites/default/files/resource-primary-download/brain.pdf

Wells, P., Tong, S., Sexton, B., Grayson, G., & Jones, E. (2008). Cohort II: a study of learner and new drivers, vol. 1. Road Safety Research Report No. 81. London: Department for Transport. URL: http://www.dft.gov.uk/rmd/project.asp?intProjectID=10067

West, R. L. (1996) An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*, 272-292.

Wester, A. E., Böcker, K. B., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis and Prevention, 40*(1), 1-7.

Wilde, G. J. S. (1994). Risk homeostasis theory and its promise for improved safety. In Trimpop, R., & Wilde, G. J. S. (Eds.), *Challenges to Accident Prevention: the issue of risk compensation behaviour* (pp. 9-24). Groningen, The Netherlands: Styx.

Williams, A. (1998). Risky driving behavior among adolescents. New perspectives on adolescent risk behavior. Cambridge: Cambridge University Press.

Williamson, A. (2003). Why are young drivers over represented in crashes? Summary of the issues Update of literature review: literature 2000 to 2003. Sydney: Report for the Motor Accidents Authority. URL: http://nla.gov.au/nla.arc-40893

Williams, A., Ferguson, S., & Shope, J. (2002). Rationale for graduated licensing and the risks it should address. *Injury Prevention, 8*(2), 9-16.

World Health Organization. Global Status Report on Road Safety 2013: Supporting a Decade of Action. Geneva: World Health Organization, Department of Violence and Injury Prevention and Disability. URL: http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/

Wyatt, J.S., Cope, M., Delpy, D.T., Wray, S., Reynolds, E.O. (1986). Quantification of cerebral oxygenation and haemodynamics in sick newborn infants by near infrared spectrophotometry. *Lancet 2*, 1063–1066.

Wylie, J. (1996). Variation in Relative Safety of Australian Drivers with Age. Report Number OR19. Canberra: Australian Publishing Service. URL: https://infrastructure.gov.au/roads/safety/publications/1996/pdf/Safe_Age_1.pdf

Yagil, D. (1998). I'm OK - you're not OK: Drivers' attitudes toward police officers enforcing traffic laws. *Policing: An International Journal of Police Strategies & Management, 21*(2), 339 – 353.

Yakovlev, P. I. & Lecours, A. R. (1967). In Minkowski, A.*, Regional Development of the Brain in Early Life (ed.)* (pp. 3–70). Oxford: Blackwell Scientific.

Yan, X., Radwan, E., & Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis & Prevention 37*(6), 983-995.

Yannis, G. (2007). Road Safety in Greece**.** Journal of IATSS, *31(2)*, 110-112.

Yoshino K., Oka N., Yamamoto K., Takahashi H., & Kato T. (2013a). Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway. *Frontiers in Human Neuroscience, 7*, 882.

Yoshino K., Oka N., Yamamoto K., Takahashi H., & Kato T. (2013b). Correlation of prefrontal cortical activation with changing vehicle speeds in actual driving: a vector-based functional near-infrared spectroscopy study. *Frontiers in Human Neuroscience, 7*, 895.

Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current Opinion in Neurobiology*, *17*(2), 251-257.

Uchiyama, Y., Ebe, K., Kozato, A., Okada, T., & Sadato, N. (2003). The neural substrates of driving at a safe distance: A functional MRI study. *Neuroscience Letters, 352*, 199–202.

Uchiyama, Y., Toyoda, H., Sakai, H., Shin, D., Ebe, K., & Sadato, N. (2012). Suppression of brain activity related to a car-following task with auditory task: An fMRI study. *Transportation Research Part F, 15*, 25–37.

UK Select Committee on Environment, Transport and Regional Affairs 19th report. (1999). Young and newly qualified drivers: standards and training. London: Stationary Office. URL: http://www.publications.parliament.uk/pa/cm199899/cmselect/cmenvtra/cmenvtra.htm

Ulleberg, P. (2001). Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. *Transportation Research Part F: Traffic Psychology and Behaviour*, *4*(4), 279- 297.

Underwood, G., Crundall, D., & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F, 14*, 435-446.

van der Zee, P., Cope, M., Arridge, S.R., Essenpreis, M., Potter, L.A., Edwards, … Delpy, D.T. (1992). Experimentally measured optical pathlengths for the adult head, calf and forearm and the head of the newborn infant as a function of inter optode spacing. *Advances in Experimental Medicine and Biology, 316*, 143–153.

Villringer, A., & Dirnagl, U. (1995) Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. *Cerebrovascular and Brain Metabolism Reviews, 7*, 240−276.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*, 748-751.

Volz, H.P., Gaser, C., Häger, F., Rzanny, R., Mentzel, H.-J., Kreitschmann-Andermahr, I., Kaiser, W.A., & Sauer, H. (1997). Brain activation during cognitive stimulation with the Wisconsin Card Sorting Test: A functional MRI study on healthy volunteers and schizophrenics. *Psychiatry Research: Neuroimaging, 75*, 145-157.

Appendices

Appendix 2.1 – Mean scores for each HP test set, for both passes and fails.

HP42

Pass/ Fail	H 003	H 135	H 142	H 165	H 188	H 229	H 250	H 290	H 474	H 521	H 570	H 735	H 747	Total
Pass	3.6	3.9	4.2	3.9	4.1	3.6	3.3	3.3	3.4	4.1	3.2	4	3.7	3.7
Fail	2.3	2.1	2.9	3	3.1	2.2	1.8	2.2	1.7	2.5	2.2	3.2	2.4	2.4

HP43.

Pass/ Fail	H 009	H 072	H 146	H 189	H 208	H 304	H 391	H 484	H 529	H 569	H 582	H 741	H 764	Total
Pass	3.3	3.3	3.1	3.7	4.4	3.4	4.1	3.9	2.8	4.2	3.9	3.3	4.3	3.7
Fail	1.7	2.3	1.9	2.7	3.4	2.1	3.1	2.3	1.5	2.7	2.9	1.8	3.2	2.4

HP44.

Pass/ Fail	H 011	H 156	H 168	H 205	H 230	H 308	H 437	H 453	H 487	H 587	H 717	H 747	H 769	Total
Pass	4.1	3.4	3.5	4	3.6	3.6	3.9	3.8	4.1	3.8	3.9	3.7	2	3.6
Fail	2.8	2.7	2.3	2.7	2.5	2.3	2.6	2.9	2.5	2.5	2.7	2.7	1	2.5

HP45

Pass/ Fail	H 030	H 065	H 139	H 148	H 176	H 213	H 317	H 394	H 498	H 508	H 510	H 591	H 715	Total
Pass	4	3.9	3.3	3.7	2.9	4.7	3.7	3.7	4.1	3.9	3.9	3.7	3.7	3.8
Fail	3.3	2.4	2.2	1.8	1.9	3.3	2.5	1.9	2.7	2.2	2.3	2.6	2.7	2.4

HP46.

Pass/Fail	H 033	H 088	H 122	H 142	H 163	H 179	H 241	H 335	H 467	H 505	H 506	H 593	H 720	Total
Pass	4.3	4.1	2.9	4.2	4	3.9	3.7	3.2	3.7	3.8	3.7	3.6	3	3.7
Fail	3.3	2.6	1.8	3	3.5	1.9	2.1	2	2.2	2.4	2.5	2.7	1.8	2.5

HP48

Pass/Fail	H 041	H 109	H 145	H 213	H 218	H 232	H 340	H 440	H 508	H 510	H 542	H 606	H 736	Total
Pass	4	4.1	3.6	4.7	3.5	4	4.1	3.6	3.9	3.9	3.7	4	2.3	3.8
Fail	3.1	2.7	2.3	3.3	2.3	2.4	2.8	2.5	2.4	2.6	1.9	2.2	1	2.4

HP49

Pass/Fail	H 003	H 135	H 142	H 165	H 188	H 229	H 250	H 290	H 474	H 521	H 570	H 735	H 747	Total
Pass	3.9	3.7	4.1	3.5	3.3	4.1	4.6	4.5	3.6	4	3.6	3.2	2.8	3.8
Fail	2	1.7	2.9	2.2	2.1	2.7	3.9	2.7	2.6	2.7	2.3	2.1	1.8	2.4

HP50

Pass/Fail	H 065	H 128	H 148	H 168	H 188	H 225	H 254	H 349	H 366	H 521	H 545	H 700	H 758	Total
Pass	4	3.7	3.8	3.5	4.3	4	3	3.3	2.9	4.1	3.5	3.8	3.8	3.7
Fail	2.6	3	2.1	2.5	3.3	2.7	1.4	2.1	2	2.8	2.2	2.3	2.3	2.4

HP51

Pass/Fail	H 055	H 088	H 113	H 162	H 183	H 228	H 286	H 434	H 498	H 528	H 569	H 635	H 748	Total
Pass	4	4.1	4.1	2.9	4	3.6	4.7	3.7	4.2	4.1	4.1	3.7	3	3.8
Fail	1.9	2.4	2.5	1.5	2.8	2.3	3.7	2	2.7	2.4	2.7	2.2	1.2	2.3

HP52

Pass/ Fail	H 107	H 158	H 180	H 229	H 290	H 401	H 529	H 558	H 569	H 700	H 716	H 747	H 764	Total
Pass	3.7	4.2	4.2	3.6	3.3	3.4	2.8	3	4.1	3.7	3.8	3.6	4.3	3.7
Fail	2.7	3.1	3.8	2.3	2.3	2.3	1.5	2	2.6	2.2	2.7	2.7	3.1	2.6

HP53

Pass/ Fail	H 109	H 118	H 156	H 159	H 215	H 230	H 258	H 366	H 395	H 421	H 537	H 584	H 741	Total
Pass	4.1	4	3.4	3.9	3.5	3.6	3.8	3	3.7	4.6	3.8	3.1	3.3	3.7
Fail	2.7	2.4	2.6	2.5	2.4	2.6	2.4	2	2.5	3	2.5	2	1.7	2.4

HP54

Pass/ Fail	H 041	H 113	H 138	H 162	H 176	H 232	H 258	H 304	H 357	H 401	H 474	H 624	H 714	Total
Pass	4	4.1	4.1	2.9	2.9	4	3.9	3.2	4.5	3.4	3.5	3.5	3.5	3.7
Fail	3.4	2.7	2.9	1.9	2.1	2.7	2.8	2	2.8	2.4	1.9	2.3	2	2.5

HP55

Pass/ Fail	H 003	H 118	H 139	H 163	H 194	H 241	H 340	H 421	H 484	H 543	H 709	H 715	H 747	Total
Pass	3.7	4	3.3	4.1	3.7	3.7	4.2	4.5	3.9	3.6	3.4	3.6	3.7	3.8
Fail	2.6	2.1	2.2	3.4	2.2	2.2	2.6	2.8	2.1	2.3	2.3	2.8	2.4	2.5

HP56

Pass/ Fail	H 033	H 120	H 129	H 164	H 245	H 273	H 304	H 434	H 487	H 545	H 716	H 720	H 749	Total
Pass	4.3	4.5	3.9	4.1	3.9	4.6	3.3	3.6	3.9	3.4	3.8	3	2.9	3.8
Fail	3.1	3.3	3	3.4	2.9	3.5	1.7	1.9	2.1	1.8	2.5	1.5	1.7	2.5

HP57

Pass/ Fail	H 011	H 122	H 130	H 165	H 211	H 254	H 308	H 510	H 552	H 595	H 624	H 717	H 736	Total
Pass	4	2.8	4.4	3.8	4.2	3.1	3.6	3.9	4.1	4.3	3.4	3.9	2.4	3.7
Fail	3	1.9	3	2.9	3	1.4	2.4	2.6	2.4	2.8	2.3	2.5	1.1	2.4

HP58

Pass/ Fail	H 040	H 072	H 126	H 135	H 189	H 205	H 258	H 317	H 498	H 529	H 553	H 718	H 748	Total
Pass	4.2	3.3	4	4	3.7	3.9	3.9	3.7	4.2	2.8	3.6	3.5	3	3.7
Fail	3.2	2.4	2.4	2.2	2.6	2.7	2.5	2.7	2.8	1.5	2.4	2.7	1.2	2.4

HP59

Pass/	H 030	H 052	H 128	H 136	H 164	H 174	H 225	H 335	H 450	H 537	H 557	H 721	H 750	Total
Pass	4	3.7	3.7	4.1	4.1	3.9	4	3.1	3.9	3.7	2.9	2.9	3.3	3.7
Fail	3.2	2.6	2.9	2.5	3.4	2.7	2.8	2	2.7	2.5	2.1	1.6	1.9	2.5

HP60

Pass/	H 002	H 129	H 145	H 180	H 225	H 273	H 349	H 453	H 528	H 560	H 714	H 722	H 769	Total
Pass	3.3	3.8	4.7	4.2	4	4.6	3.5	3.8	4.2	3.4	3.5	3.3	1.9	3.6
Fail	2.1	3.1	2.5	3.9	2.7	3.8	2.3	2.9	2.8	1.9	2	2.1	1	2.6

HP61

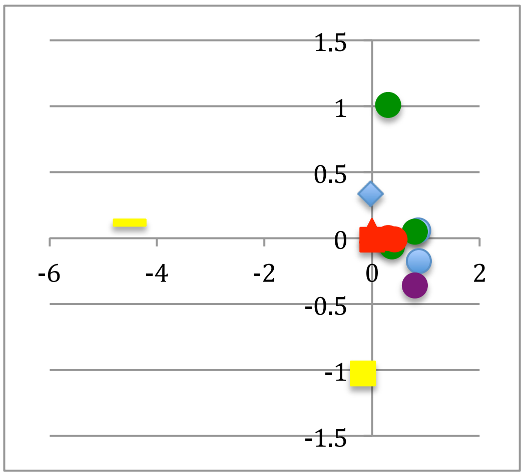
Pass/	H 002	H 052	H 120	H 126	H 167	H 183	H 192	H 208	H 467	H 505	H 561	H 705	H 754	Total
Pass	3.2	3.7	4.5	4.1	4.1	4.1	3.4	4.4	3.7	3.9	3.6	2.6	3.1	3.7
Fail	2.1	2.7	3.4	2.4	2.7	3.1	2.1	3.2	2.2	2.5	2	1.5	1.2	2.4

HP62

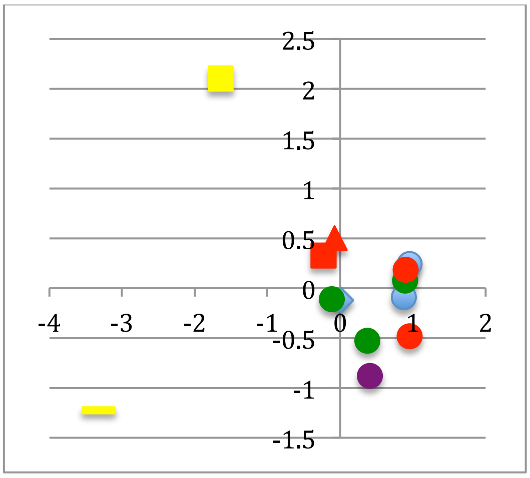
Pass/	H 107	H 143	H 158	H 174	H 192	H 211	H 245	H 337	H 453	H 505	H 584	H 624	H 635	Total
Pass	3.6	4.2	4.1	3.8	3.3	4.2	3.8	3.6	3.7	3.8	3.2	3.4	3.7	3.7
Fail	2.6	2.8	3	2.7	2.2	3	3.1	1.9	2.9	2.4	1.9	1.9	2.3	2.5

Appendix 2.2. MDS scaling output for the reamining HP test sets, for both passes and fails.

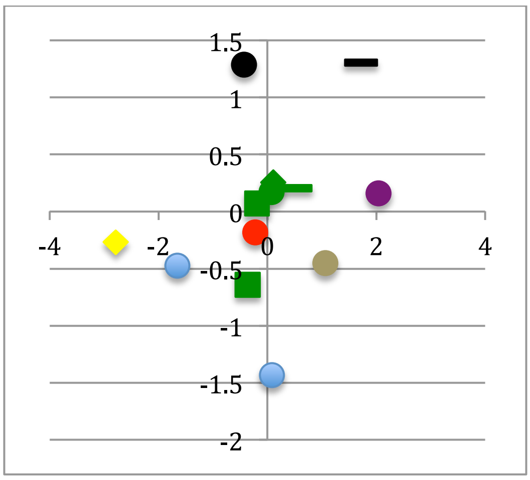
HP44 Pass



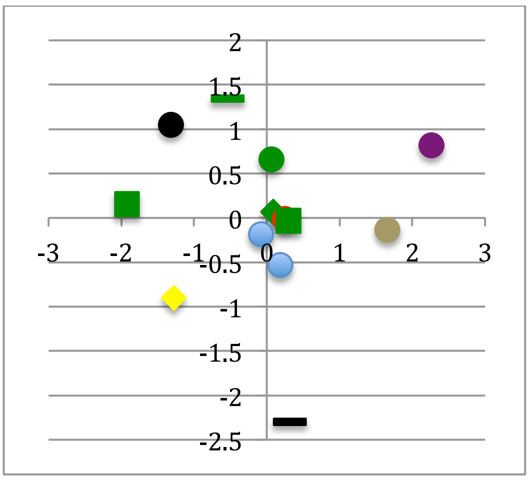
HP44 Fail



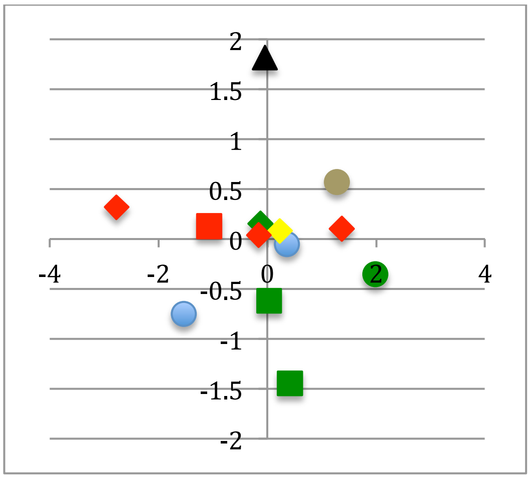
HP45 Pass



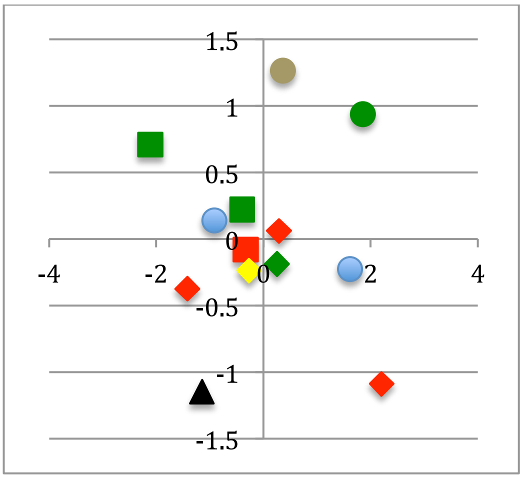
HP45 Fail



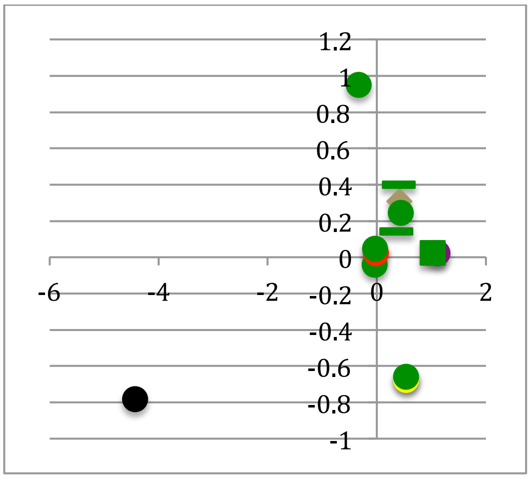
HP46 Pass



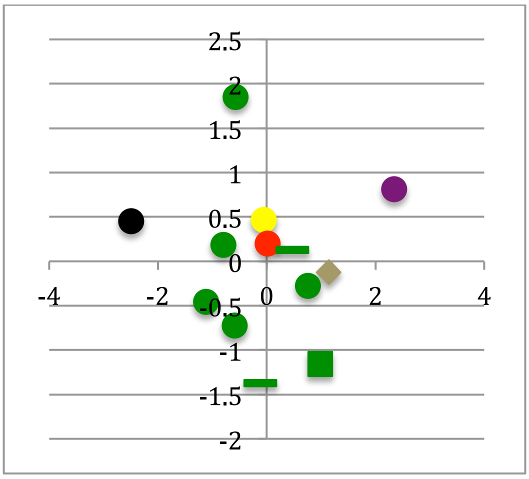
HP46 Fail



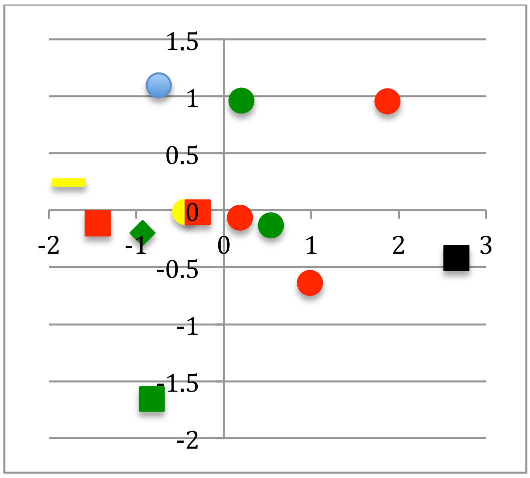
HP48 Pass



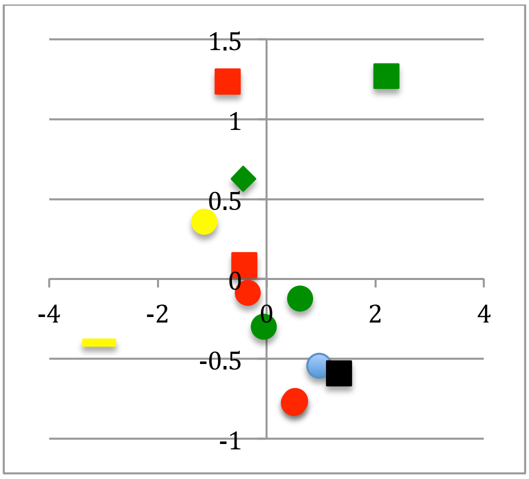
HP48 Fail



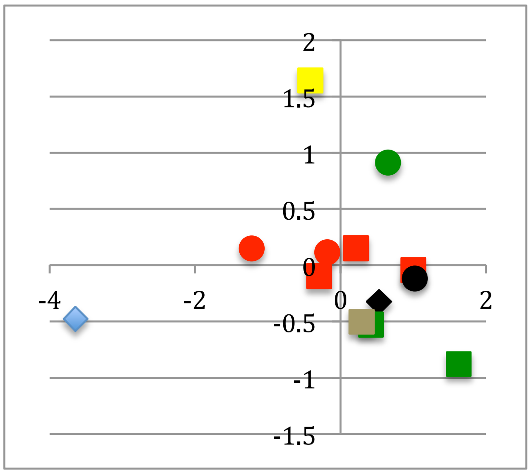
HP49 Pass



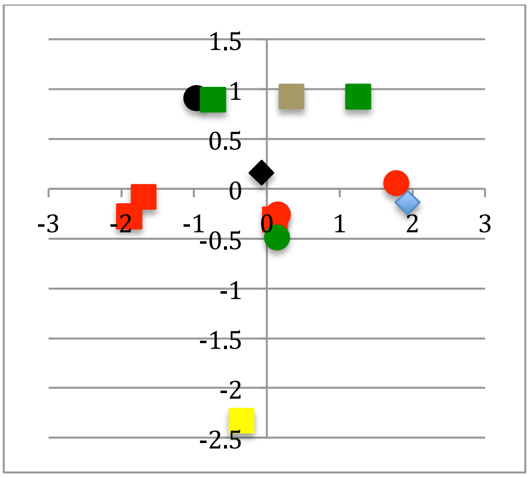
HP49 Fail



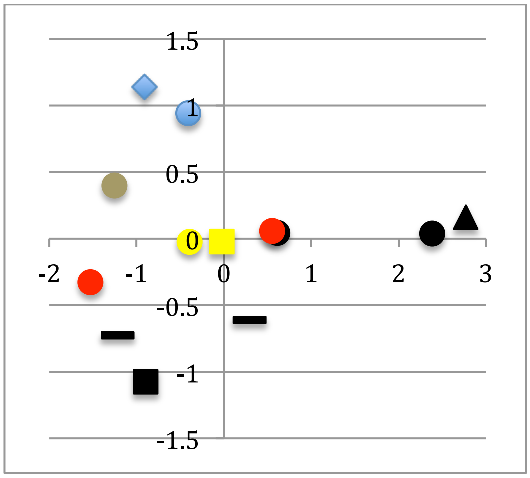
HP50 Pass



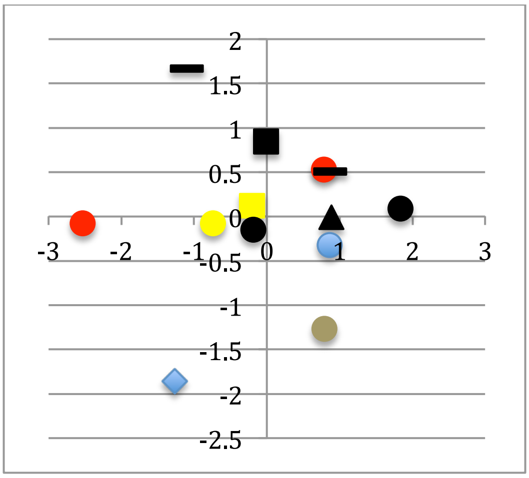
HP50 Fail



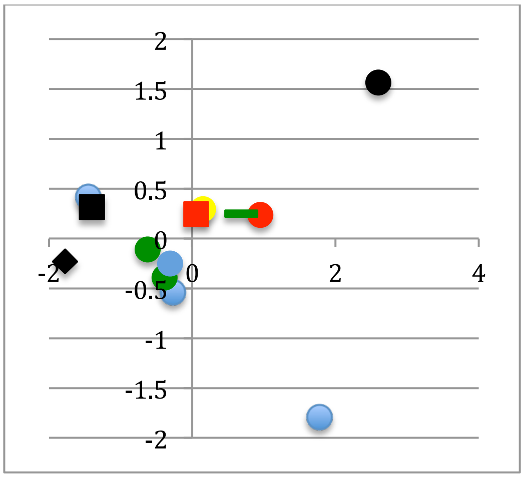
HP51 Pass



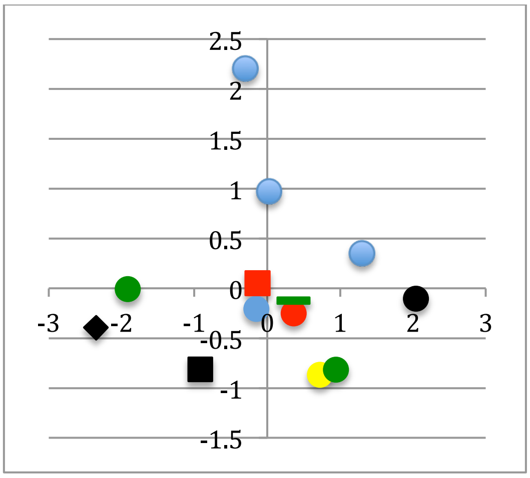
HP51 Fail



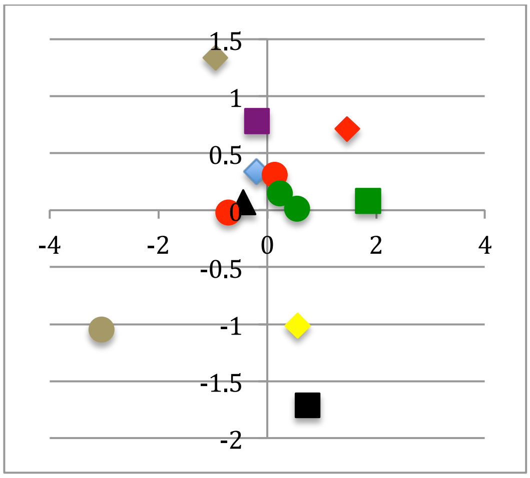
HP52 Pass



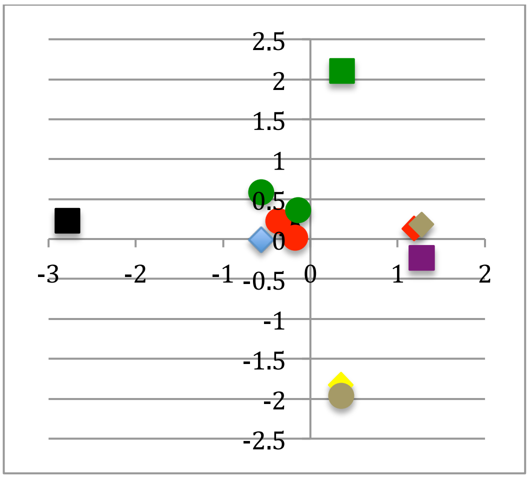
HP52 Fail



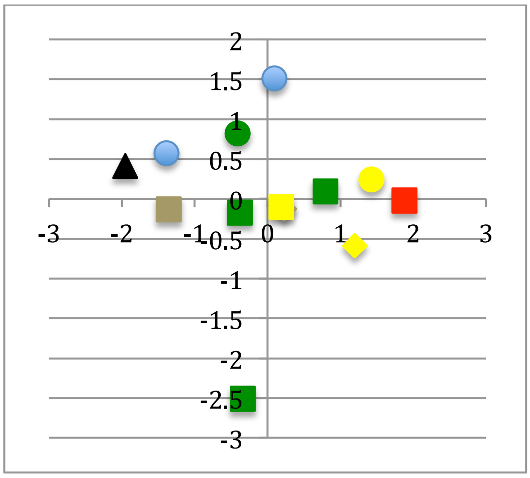
HP53 Pass



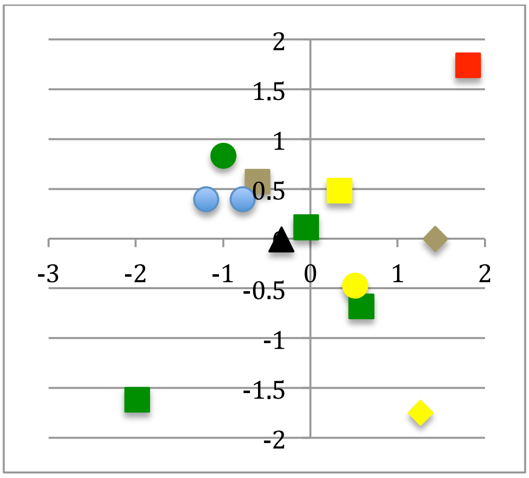
HP53 Fail



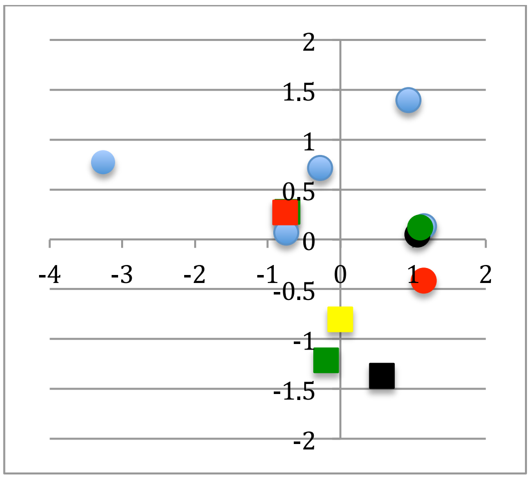
HP54 Pass



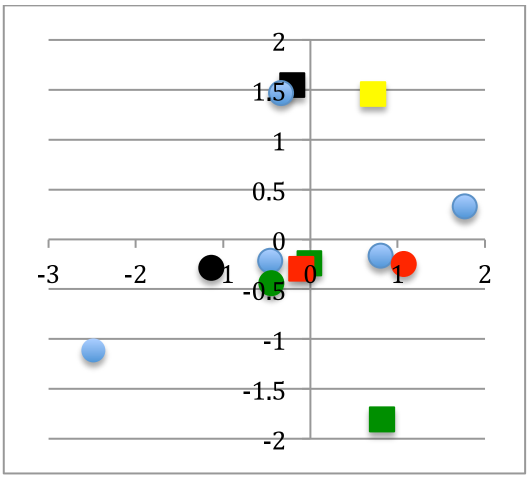
HP54 Fail



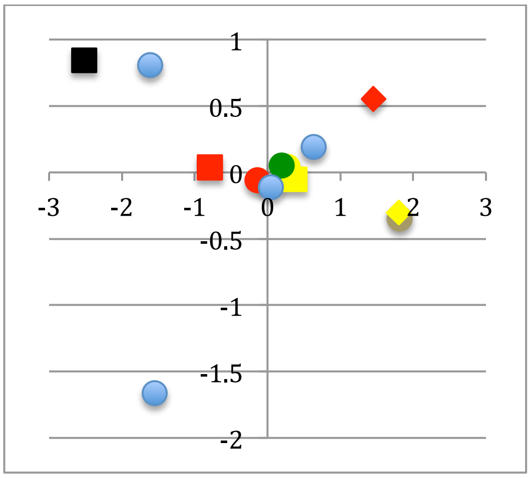
HP55 Pass



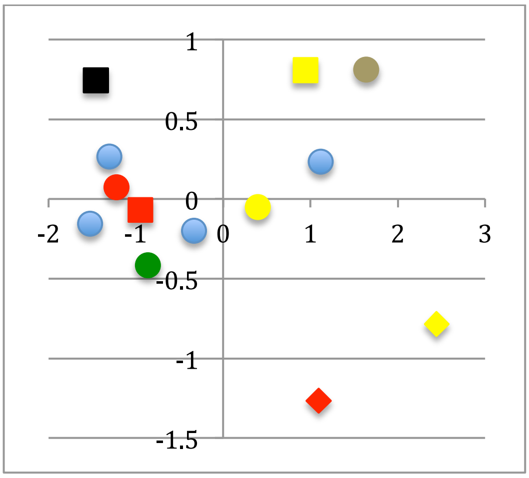
HP55 Fail



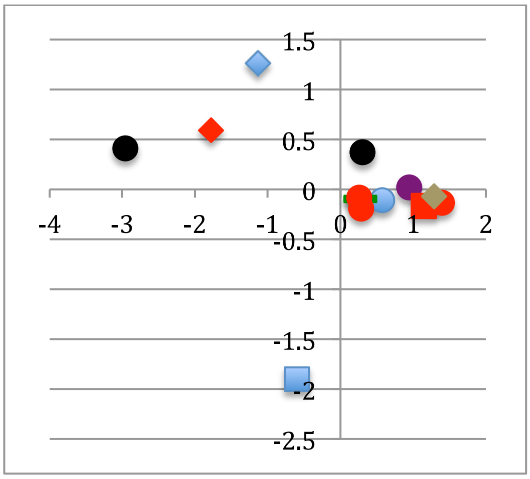
HP56 Pass



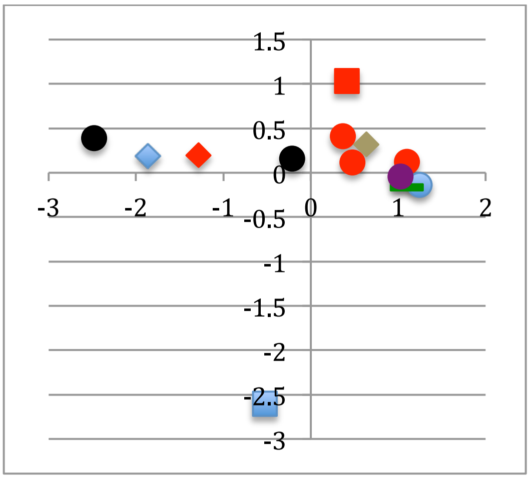
HP56 Fail



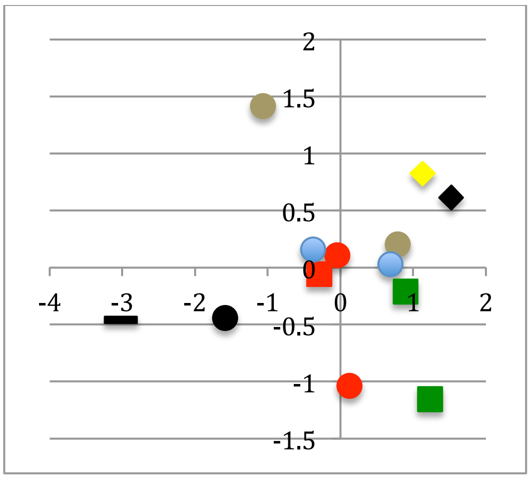
HP57 Pass



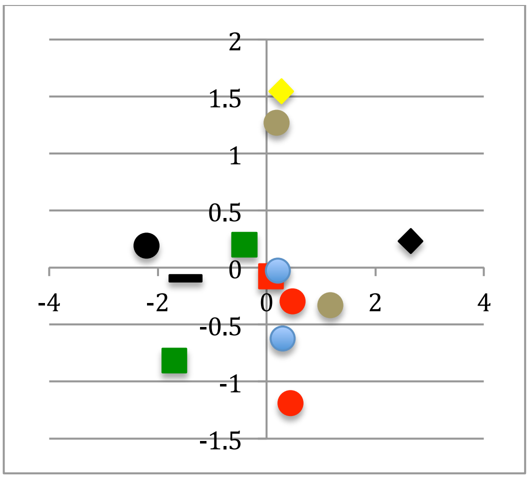
HP57 Fail



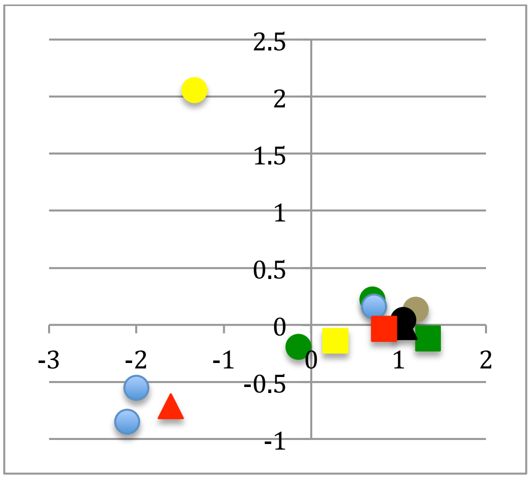
HP58 Pass



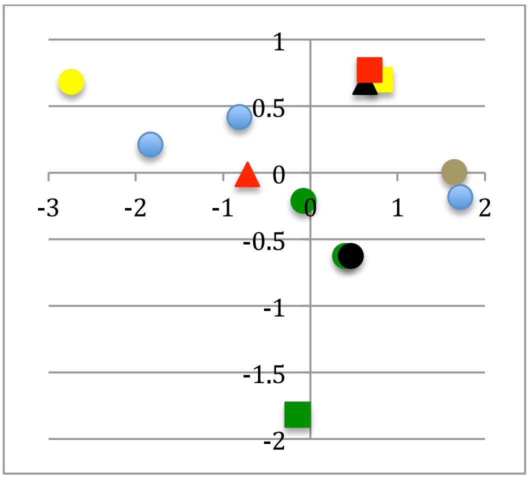
HP58 Fail



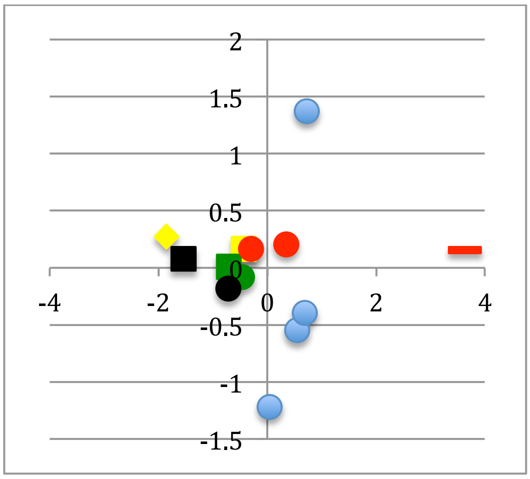
HP59 Pass



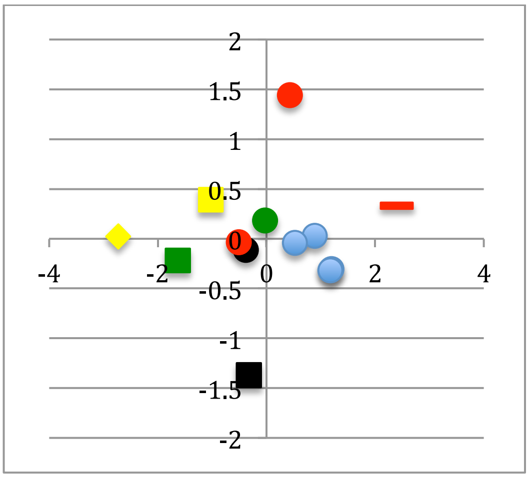
HP59 Fail



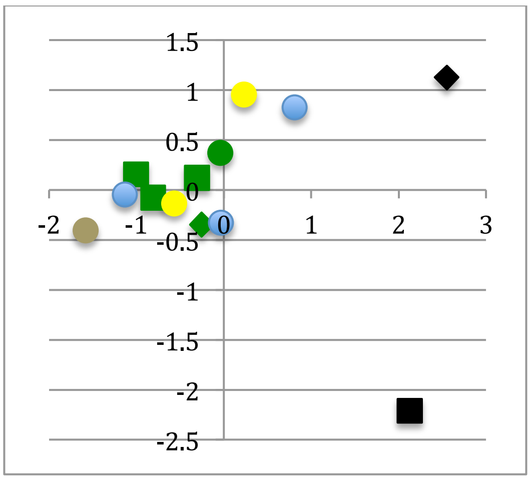
HP60 Pass



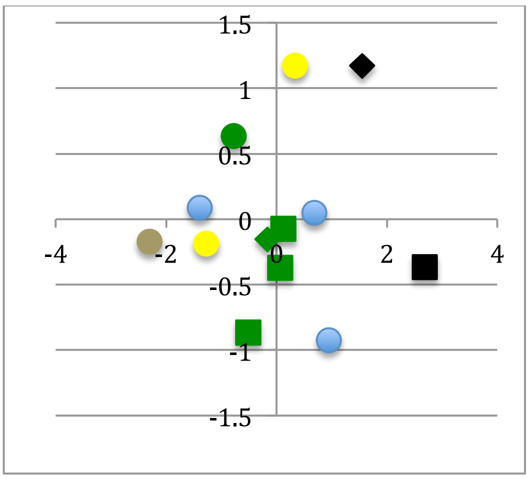
HP60 Fail



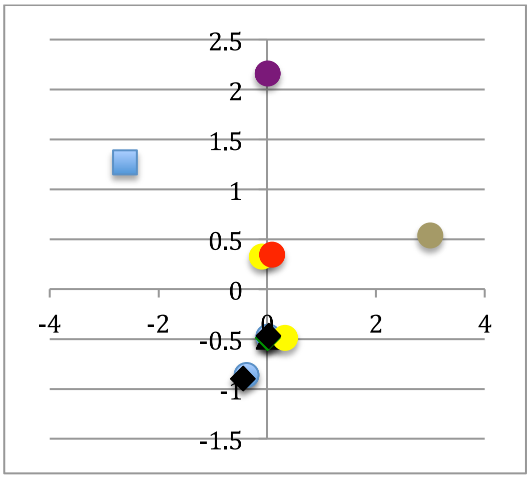
HP61 Pass



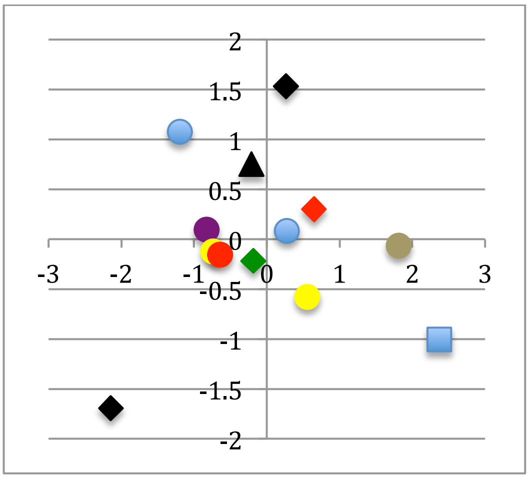
HP61 Fail



HP62 Pass



HP62 Fail



Appendix 3.1 Simulator sickness questionnaire

Please tick the one that applies for each symptom.

	None	Slight	Moderate	Severe
1. General discomfort				
2. Fatigue				
3. Headache				
4. Eyestrain				
5. Difficulty focusing				
6. Salivation increased				
7. Sweating				
8. Nausea				
9. Difficulty concentrating				
10. "Fullness of the head"				
11. Blurred vision				
12. Dizziness eyes open				
13. Dizziness eyes closed				
14. Vertigo				
15. Stomach awareness				
16. Burping				

Appendix 5.1 General Instructions

Welcome to the experiment and thank you for taking the time to help us with this research. Throughout the course of this experiment you will be asked to take part in three separate tasks whilst wearing a piece of equipment called a functional near-infrared spectroscopy (fNIR) device. This piece of equipment is used for brain imaging, with this specific device used for measuring haemodynamic changes in the frontal lobes. The device works by passing light through your forehead and identifying the amount of oxygenated and deoxygenated haemoglobin (which transport oxygen in red blood cells) in the frontal lobes.

The fNIR device will be placed on your forehead when you are sat facing the computer screen and will remain on for the duration of the three tasks, which have been specifically chosen as they have been shown to be dependent on the frontal lobes. It is important that during the three tasks you keep your head as still as possible.

If you are happy with these instructions given to you so far and you are still happy to take part in this experiment, please let the experimenter that you would like to continue. If you have any questions about any aspect of the experiment then please ask the experimenter.

Appendix 5.2 WCST Instructions

There are two parts to this experiment.

<u>PART 1</u>

4 different reference cards will appear together at the top of screen adjacent to one another on every trial. On every trial these reference cards will remain the same.

Each reference card has a corresponding button on the keypad in front of you. You are to use you left index and middle finger for the 2 left orange buttons (corresponding to the 2 left cards respectively) and your right index and middle fingers for the 2 right orange buttons (corresponding to the 2 right cards respectively).

These reference cards and button correspondents will remain the same throughout the entire experiment.

Take a moment to familiarise yourself with the buttons in front of you.

In addition to the reference cards, there will be a stimulus card centred at the bottom of the screen. This part of the experiment will require you to either:

- Simply match the stimulus card to one of the reference cards, as they will be exact replicas of each other
- Deduce via trial and error as to which single sorting criterion the stimulus card, which will differ in some way to the reference cards, associates with the reference cards

The sorting criterion will not always be the same and may change throughout the experiment. You will be given onscreen feedback each time stating that your choice was either 'CORRECT' or 'INCORRECT'.

Part 2

As in part 1, the same 4 reference cards will appear each time, with the same corresponding buttons to respond with. The aim will be to match a stimulus card to one of the 4 reference cards based on a matching criterion.

This time the criterion will be given to you every set of trials. You simply have to abide by the criterion rule given until the next criterion is provided.

If you understand the rules and are ready to begin, please notify the experimenter.

Appendix 5.3 CBT Instructions

A series of 9 blocks will appear on screen. One target block will light up RED every second and will be in sequences of 3, 4, 5 or 6 target blocks. Please remember the sequence of these blocks in the correct order. You will be required to recall this sequence.

On some trials a block will light up in yellow alongside the red block. It is important to remember the sequence of the RED block.

After the sequence is shown, you will be required to recall the sequence in the correct order. You will be given two response options between two blocks and you will have to make a decision whether one of these blocks was part of the original sequence.

These two block options are labelled with 'L', which indicates a left response and 'R', which indicates a right response.

Please respond as accurately and quickly as possible, based on whether you had seen that block in the previous sequence in the correct order.

Appendix 5.4 COWAT Instructions

Generation Condition

You need to think of as many words as you can that begin with a particular letter in 1 minute.

Speak them aloud so that the experimenter can record the words that you think of.

Try to avoid using homophones (words that sound the same but have different meanings) and extensions (words you have already used with just minor changes).

Tell the experimenter when you are ready and the then the first letter will be displayed.

Control Condition

You will now be presented with a fixation cross, followed by a word.

Simply read each word aloud as it is presented.

Appendix 5.5 COWAT reading condition word list.

Letter	F	A	R	S
	Food	Awful	Red	Shake
	Fast	Awake	Rack	Ship
	Fun	Amazing	Roll	Snow
	Face	Acrid	Run	Student
	Fin	Amnesia	Rush	Soup
	Fantastic	After	Rucksack	Sun
	Fish	Almighty	Ranch	Show
	Frog	Apple	Rick	Spiders
	Freedom	Alight	Range	Soldiers
	Feel	A lot	Recoil	Spelling
	Forgiveness	Almond	Repeat	Sham
	Floor	Aware	Round	Search

Appendix 6.1 NASA-TLX Scale

NASA-TLX Mental Workload Rating Scale

Please place an "X" along each scale at the point that best indicates your experience with the display configuration.

Mental Demand: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc)? Was the mission easy or demanding, simple or complex, exacting or forgiving?

Low High

Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the mission easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

Temporal Demand: How much time pressure did you feel due to the rate or pace at which the mission occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

Performance: How successful do you think you were in accomplishing the goals of the mission? How satisfied were you with your performance in accomplishing these goals?

Low High

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

Frustration: How discouraged, stressed, irritated, and annoyed versus gratified, relaxed, content, and complacent did you feel during your mission?

Low High