INVESTIGATION OF POLICE DECISION MAKING USING COMBINED EEG AND VIRTUAL REALITY METHODS

NICHOLAS ANDREW ALEXANDER

Doctor of Philosophy

ASTON UNIVERSITY

June 2020

Aston University

Investigation of Police Decision Making Using Combined EEG and Virtual Reality Methods

Nicholas Andrew Alexander

Doctor of Philosophy

2020

Police officers in the UK are granted additional powers to allow them to protect life and property and prevent crime. Of these powers, the sanction to use stopping, potentially lethal, force given to Authorised Firearms Officers (AFO) is arguably the most salient. Each decision made by an AFO to discharge their firearm or not has great impact and so it is important we research the cognitive processes that lead to such a decision.

One challenge of researching these cognitive processes is eliciting ecologically valid behaviour while maintaining internal validity. We approached this challenge by developing combined electroencephalography (EEG) and virtual reality research methods. Using these methods, we produced scenarios that reflected features of AFO training. First, we tested simple versions of the scenarios on a novice population. Following this, we increased the complexity of the scenarios and collected data from both AFOs and novices.

We found that participants were fastest when responding to threatening scenarios. Further, AFOs had consistently faster response times than novices. In line with similar 'Go/No-Go' paradigms, we found greater increases in pre-response frontal-midline theta when participants did not shoot versus when they did. Comparisons of EEG between AFOs and novices revealed greater pre-response increases in frontal-midline theta and central delta when they equipped a firearm. Greater differences in delta activity were also observed between different levels of threat in the AFO group.

Together, these findings suggest that differences in performance between experts and novices may be due to their greater attention towards threat. Further investigation of expert decision making should build on our use of naturalistic stimuli and expert participants to ensure that findings are ecologically valid. With increasing accessibility of modern game engines and virtual reality technology, this approach will be beneficial to researchers in many fields where ecological validity is required.

Keywords:

Electroencephalography; Expertise; Shoot/don't-shoot; Head-mounted display; Naturalistic imaging.

Acknowledgements

First and foremost, I would like to thank my academic supervisors, Prof. Klaus Kessler and Dr Robert Nash at Aston University and Prof. Matt Brookes at the University of Nottingham, for their guidance and support throughout our project. Klaus, thank you in particular for the tremendous amount of time and energy you have invested in our project and into my academic and personal development.

This project has been a collaboration involving four organisations: Aston University, the University of Nottingham, Durham Constabulary and Cleveland Police. At the heart of this collaboration was Shaun Beebe, who I would like to thank not only for his contributions and unique perspective regarding this project, but also for the exceptional personal and professional advice he has given me.

I would like to thank Aston University and the University of Nottingham for funding my studies and all aspects of this project. Thank you to all members of the ALIVE and EXES groups for making it such an enjoyable place to work. In addition, thank you to Hongfang Wang for guiding me through EEG analysis and programming, and to Clíona Kelly and Nafeesa Gul for recruiting and testing participants.

Thank you also to Durham Constabulary and Cleveland Police, who enabled this whole project. I would like to thank Former Chief Constable Mike Barton for backing the project, Andy Sutherland and Karl Brown for their insight on policing, and Michaela Bourke for organising my visits.

Participants in our experiments included staff and students at both Aston University and the University of Nottingham, members of the public and Authorised Firearms Officers within Durham Constabulary and Cleveland Police. Thank you to all for their time and interest.

Everyone above and more has contributed directly to the work presented in this thesis. My own contribution has only been possible due to the fantastic support network around me. Fortunately, there are too many to name, but thank you to all the wonderful friends I have made while studying and to all the old ones who have visited me in Birmingham. To mention one, thank you Jack for putting me up while I finished my thesis. Thank you, Mum, Dad, Jen, Dave and Oli, and all my family, for their unconditional support. Above all, thank you Joanne for your love, sacrifice and commitment throughout. I dedicate this thesis to you.

List of Contents

List of Abbreviations

3D	Three-dimensional
6DOF	Six degrees of freedom
ACC	Anterior cingulate cortex
ACTH	Adrenocorticotropic hormone
AFO	Authorised Firearms Officer
BART	Balloon Analogue Risk Task
BEM	Boundary element method
CAVE	Cave automated virtual environment
CI	Confidence interval
CRF	Corticotrophin-releasing factor
CSD	Cross-spectral density
df	Degrees of freedom
DRT	Decision response time
EEG	Electroencephalography
eLORETA	Exact low-resolution brain electromagnetic tomography
ERN	Error-related negativity
ERP	Event related potential
EU	Expected Utility
FBI	Full Body Illusion
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near infrared spectroscopy
GDT	Game of Dice Task
HMD	Head-mounted display
HPA-axis	hypothalamic-pituitary-adrenal axis
Hz	Hertz
IGT	Iowa Gambling Task
kΩ	Kiloohm
M	Mean

m	Metre
MN	Matched Novices
ms	Millisecond
MTPL	Motion-to-photon latency
NDM	Naturalistic Decision Making
Neuro-VR	Neuroimaging and virtual reality
OP-MEG	Optically pumped magnetometer-based magnetoencephalography
PFC	Pre-frontal cortex
PRT	Preparation response time
PSD	Power spectral density
RHI	Rubber Hand Illusion
RPD	Recognition-Primed Decision
s	Seconds
SAM system	Sympathetic adrenomedullary system
SD	Standard deviation
SEP	Somatosensory evoked potential
SEU	Subjective Expected Utility
SMH	Somatic Marker Hypothesis
SQUID-MEG	Superconducting quantum interference device-based magnetoencephalography
UE4	Unreal Engine 4
vmPFC	Ventral-medial prefrontal cortex
YN	Younger Novices
η_{ges}^2	Generalised eta squared

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Aims and rationale

We rely upon experts in many fields to make optimal decisions consistently, with limited information, and often in highly adverse situations. The effectiveness of professionals in industries like the emergency services is a testament to experts' ability to do so. It is without doubt that, when given the same task, novices who lack experience and training, and who have not been through the same selection processes as these experts, would perform far less effectively. In this research project, we aimed to use the difference in performance between experts and novices as a manipulation with which to investigate decision making. We focused our investigation on Authorised Firearms Officers (AFOs) in the UK Police Force. It is of particular importance that we understand the processes that lead to the decisions that AFOs make because, while they are civilians, they are granted extraordinary powers so that they may protect life and property and prevent crime (College of Policing, 2014). In particular, they can be authorised, and ultimately self-authorise, to discharge a firearm within the bounds of domestic and international law (National Policing Improvement Agency, 2011). When an AFO discharges a firearm, they intend to cause stopping force, but the result may be lethal, and, as civilians, each officer is accountable for their own actions (College of Policing, 2013). Further, because of the powers they have, the decisions made by AFOs are rigorously scrutinised–not only within the police and legal system, but in the news and media as well (Rogers, 2003). The result of this scrutiny greatly affects public opinion of the police and this in itself has wide-ranging impact (Hough et al., 2010). Crucially for this project, the study of AFOs' decision making is more feasible than for some other groups as the behaviour we want to measure occurs in a single, clearly defined moment: the pulling of the trigger.

1.2. Neuro-VR approach

One of the first challenges of the project was to understand how best to create a paradigm that could allow our measures to be sensitive to the specific behaviour of AFOs. Hope (2016) outlined some methodological standards that should be met when conducting research on police performance which we used as guidance. Alongside good scientific practice, Hope (2016) recommended the use of: (1) experimental manipulations relevant to policing (i.e. not treating police officers as a general 'special case' in tasks outside their own expertise); (2) realistic, controlled scenarios; and (3) control groups.

These recommendations were based on common limitations observed in research on police officer performance and we wanted to avoid those pitfalls.

To allow for suitable comparison against a control group, one requirement was that both novices and expert AFOs must be able to complete the task. This ruled out anything that required firearms training or qualification. For example, we could not compare the performance of experts and novices during AFO training, and certainly not during their response to real incidents. For this reason, expert decision making in the field is more often evaluated qualitatively and without direct comparison to novices (e.g. Harris et al., 2017; Klein, 2008). Fortunately, when this project was started, circa 2017, the first generation of modern virtual reality head-mounted displays were becoming available. While capable virtual reality technology has existed for some time, this new generation represented a step change in affordability and accessibility to researchers in fields outside of neuroscience and industry, as the products were available at a consumer level (Slater, 2018). One immediate benefit of virtual reality to research is that it makes "the impossible possible" (de la Rosa & Breidt, 2018). For our purposes, this meant we could create simulations that approximated the training or real-world situations that AFOs make decisions in. Crucially, these simulations could be completed without expertise, so direct comparisons between AFOs and novices could be made. To develop this concept into an experimental paradigm, we formed a collaboration with the Durham Constabulary and Cleveland Police; through their shared Tactical Training Centre, we worked with senior firearms instructors who create and deliver the training for AFOs. They shared their knowledge about the decision-making process instilled throughout AFO training and outlined common scenarios and a range of appropriate decisions AFOs might make.

To make these scenarios and use them as trials within a neuroscience experiment, we used tools readily available in the entertainment and game development industries for the creation and presentation of stimuli. In doing so, we followed a common trend in science: to take technologies developed for other applications and employ them as research methods. In no field is this more true than neuroscience, which has seen great success from using advances in imaging techniques to record different aspects of the metabolic processes that cause human behaviour (Bandettini, 2009). We can observe this trend elsewhere in the field: for data analysis, neuroscientists have worked rapidly to integrate new developments in signal detection theory, machine learning and statistics to take full advantage of the benefits they offer for working with complex datasets (e.g. Friston et al., 1994; Haxby et al., 2001; Marblestone et al., 2016). Broadly speaking, computers and the internet have changed the way that data is collected, and technologies such as eye tracking and motion capture have made these data richer (e.g. Bridges et al., 2020; Popa et al., 2015). Despite this eagerness to use new methodologies, when it comes to modern tools for stimulus development that exist outside of neuroscience, researchers have been more resistant to change (Sonkusare et al., 2019). Because of this, using these tools for our project added the challenge of integrating them with established cognitive neuroscience methods. Therefore, the second aim of our project was to develop and validate these novel methods–not just for use in this project, but for other research projects that would benefit from using them. We refer to this integration as 'neuro-VR'.

While neuro-VR could employ a range of neuroimaging techniques, some are more practical than others. The technology used to present virtual reality contains many electronics and sensors (Anthes et al., 2016; Cruz-Neira et al., 1992; Lavalle et al., 2014), and so techniques which rely on the generation and/or observation of magnetic fields, such a functional magnetic resonance imaging (fMRI) and superconducting quantum interference device-based magnetoencephalography (SQUID-MEG) are currently impractical, as they would interrupt/be interrupted by them (Dempsey et al., 2002; G. Roberts et al., 2019). In addition, imaging techniques which prohibit movement of the head and body, which again includes fMRI and SQUID-MEG (Medvedovsky et al., 2007; Qin et al., 2009), are less suitable for neuro-VR (G. Roberts et al., 2019). The reasons for the importance of movement and other actions to be possible in virtual reality will be discussed in more detail in Chapter 2. After these exclusions, there are at least three viable options for neuro-VR: electroencephalography (EEG); functional near infrared spectroscopy (fNIRS); and optically pumped magnetometer-based magnetoencephalography (OP-MEG). Arguably the most mobile neuroimaging technique available today is EEG (Jungnickel et al., 2019). The amplifiers, computer and storage required for EEG are now able to fit inside a small backpack and can produce recordings of comparable quality to traditional wired EEG (Ries et al., 2014). Modern data acquisition and signal processing techniques allow recordings to be made in more natural settings with more natural movement, without reducing the quality of the data. For example, EEG can be recorded during running and walking (Gramann et al., 2014; Nathan & Contreras-Vidal, 2016). Progress towards mobility is being made in other technologies, such as OP-MEG, which allow increased range of movement and less susceptibility to environmental noise than SQUID-MEG (Barry et al., 2019; Boto et al., 2018). fNIRS is also particularly well suited for neuro-VR and has benefited from similar technological improvements as EEG, along with miniaturisation of sensors, allowing for greater range of movements required for virtual reality (Brigadoi et al., 2019; Quaresima & Ferrari, 2019). Unfortunately, even compact neuroimaging methods that do not prohibit head movement work best when it is limited. For EEG, OP-MEG and fNIRS, sensors must remain fixed to the wearer's scalp. While this allows for some head movement, all of these methods suffer from movement artefacts introduced when head movement causes small changes in their position, relative to the scalp: EEG requires constant contact between electrodes and the scalp and fNIRS and OP-MEG assume a constant position of sensors relative to the scalp (Hari & Puce, 2017; Hill et al., 2020; Quaresima & Ferrari, 2019). Further, EEG and, to a lesser extent, OP-MEG, are both susceptible to noise caused by muscle activity proximate to sensors (Boto et al., 2018; Muthukumaraswamy, 2013; Zimmermann & Scharein, 2004). fNIRS is also sensitive to changes caused by muscles close to sensors (Schecklmann et al., 2017), and the effects of changes in blood pressure in areas near the sensors caused by movement and exertion are also a problem for analysis (Molavi & Dumont, 2012). Nonetheless, virtual reality has been successfully combined with EEG (e.g. Bayliss & Ballard, 2000; Tromp et al., 2018), fNIRS (e.g. Landowska, Roberts, et al., 2018; Landowska, Royle, et al., 2018), and OP-MEG (G. Roberts et al., 2019), and so these methods remain the most feasible candidates for neuro-VR research.

1.3. Why EEG?

We elected to use EEG in our investigations of AFO decision making. This was because, alongside its suitability for neuro-VR, EEG has high temporal resolution (in the order of milliseconds) which offers advantages over other mobile functional neuroimaging methods, such as fNIRS, when investigating dynamic brain function (Lenartowicz & Poldrack, 2010). We also acknowledged the limitations of EEG. Primarily, the lower spatial resolution of EEG compared to fNIRS and, even more so, OP-MEG (Boto et al., 2018). This would limit any conclusions we could draw about where in the brain any differences our analysis found were. Rather than rely of source localisation to differentiate experimental conditions, we compared how known signatures of brain activity varied over time. Within the limits offered by EEG, we could then use source analysis to support identification of these signatures.

Recording EEG involves measuring changes in electrical potential between multiple positions on the scalp using highly sensitive electrodes and voltmeters (Michel & Murray, 2012). As such, it is a direct measure of neuronal activity. There are many sources of electrical activity which can cause changes in the electrical potential recorded by these electrodes (Teplan, 2002). EEG refers to those changes that arise from metabolic processes in the brain. Briefly, the membrane potential of neurons can change as they exchange ions to maintain resting potential or to generate action potentials. Synchronous change in the membrane potential of many spatially aligned and proximate neurons can generate enough electrical current to be detected at the scalp following volume conduction (Hari & Puce, 2017; van den Broek et al., 1998). Following the assumption that these spatially aligned and proximate neurons are involved in a shared function, their activity is representative of that function. These functions may include sensory perception or cognitive processes. By recording EEG during known behaviours or during the presentation of stimuli, cognitive processes specific to events can be identified.

Two common phenomena emerge when analysing event-locked EEG: event related potentials (ERP) and changes in power constrained within frequency bands (Hari & Puce, 2017). These frequency bands include delta (1-4Hz), theta (4-7Hz), alpha (8-12Hz), beta (15-25Hz) and gamma (>30Hz). ERPs are waveforms identified by averaging many event-locked EEG recordings. They can be said to be event related, because any change not related to the event (i.e. noise) would be averaged out. Using this method, the exact cause of an ERP cannot be known. This is because there are two kinds of event related change in EEG: induced change, where ongoing activity is modulated, and evoked change, where new activity is produced (Pfurtscheller & Lopes da Silva, 1999). It is commonly assumed that the waveform of an ERP represents evoked change in response to an event. The reason for this assumption is that if induced change is not phase-locked then it would be averaged out with the noise. This is not necessarily the case because induced change, such as event related phase realignment, could also produce ERP waveforms (Burgess, 2012). However, it is certainly the case that many induced changes in activity are not detected as ERPs (Pfurtscheller & Lopes da Silva, 1999). To measure these, analysis must move away from the time domain and consider how power changes in different frequencies over time. By doing this, two benefits are gained. First, non- phase-locked changes in activity can be detected. Second, and crucial for any research with less well-defined event timing, such as that presented in this thesis, analysis in the time-frequency domain usually considers a wider epoch and is not as affected by jitter.

1.4. Overview of thesis

In this thesis I will present research working towards our aims, culminating in an experiment comparing expertise between AFOs and novice control groups. In addition to contributing to the investigation of expertise more generally, our findings inform on the specific decision-making processes that AFOs use. We continue to work with our partners in the police with the hope that we can contribute to the knowledge which goes into AFO training at facilities such as the Tactical Training Centre.

Chapter 2: Literature Review

2.1. Guide to review

The central theme of this literature review is the measurement of naturalistic decision making in virtual reality using EEG, with a focus on police decision making. As this is a multifaceted, but ultimately narrow research area, I will start by reviewing the literature in each of its constituent components: decision making, naturalistic stimuli, virtual reality, and police performance. I will work towards combining these components into more focused areas of naturalistic decision making and neuro-VR. Throughout my review I will attempt to highlight common themes across fields and how they relate to our empirical work.

Some of these topics could be very broad because the same terms are used in different ways–so much so that a true review of 'decision making' would include research on a variety of topics, such as single cell organisms (Dexter et al., 2019) or the study of machines (Lake et al., 2017). Likewise, while 'expertise' is generally used to describe increased performance from training or experience, the extent of it varies. Therefore, it is important to operationalise these terms now. Decision making will only refer to those decisions made by individual humans. Expertise will refer only to changes in performance due to substantial training and experience, rather than brief laboratory manipulations. Later, the definition of virtual reality will be discussed in detail, with critical reference to its use in the literature.

2.2. Decision making

2.2.1. Classical Decision Making

The problem that a decision must solve is the choice between multiple actions. The way in which this problem can be approached is determined by the availability of information (Ellsberg, 1961). When information is known about the probabilities and outcomes of each action, a decision can be said to be made under risk or certainty (in the case of only one outcome). One way of phrasing these decisions is as a choice between lotteries, where each lottery contains a set of outcomes and probabilities (Schoemaker, 1982). Value can be assigned to each outcome according to some interpretation of its utility. Normative models of decision making, which seek to describe the norms for how people should act, can apply probability theory to predict the optimal choice (Marschak, 1950). In their most crude form, normative models could describe probability and utility as linear, or having cardinal properties. This would mean that decisions are made by solving independent probability problems for each action and comparing the result along a single scale of utility.

von Neumann & Morgenstern (1944) produced a series of axioms, or rules, which described decision making under risk in terms of Expected Utility (EU). Briefly, EU for each action can be measured with reference to a utility function and then ranked (for a review, see Schoemaker, 1982). Importantly, the EU of each action falls along the same linear scale. This means every action's EU can be compared independently, without the need for a mediator, as defined under the axiom of transitivity (Karni, 2014). The EU model works very well for the normative approach because it describes what action should be taken by an individual when making a risky decision. However, even in fields like economics, where vast amounts of data are collected and tested, and certainly for most of the decisions that we make as individuals, it is unlikely that many decisions are truly made under risk, let alone certainty; outcomes and their probabilities are rarely so well defined (Ellsberg, 1961). More often, decisions are made under uncertainty. Here, uncertainty describes the situation when, for a given action, all outcomes are known but their probabilities are not. Fishburn (1989) and Savage (1954) developed Subjective Expected Utility (SEU) theory to tackle this problem. This more flexible approach reduces cardinal assessment of probabilities to ordinal, more qualitative descriptions. In SEU theory, probabilities can be ranked in relation to each other, without common factors or a shared utility scale. Fishburn's (1982) axiomatisation of SEU necessarily challenges some of the foundations of the EU model. For example, SEU theory does not allow for transitivity, because not all actions' utility can be compared. Like EU, SEU theory is a useful normative model for decision analysts in many fields, including medicine and economics (Luce & von Winterfeldt, 1994). Specifically, it is useful for understanding how best to act under uncertainty.

Unfortunately, while the strict rationality of both EU and SEU theories state the norms for how a decision maker should act, they do not adequately describe the behaviour of human decision makers (Marschak, 1950). The first piece of evidence for this comes from apparent differences both within- and between-individuals. If everyone followed a rational norm, then only one choice per decision task would ever be made. Rather than try and explain all this variation, a model should attempt to explain "the behaviour of men [and women] who, it is believed, cannot be “all fools all the time”” (Marschak, 1950, p. 111). It has been consistently demonstrated that humans systematically underperform against normative model predictions when presented with risk or uncertainty (Camerer & Weber, 1992; Weber & Camerer, 1987). Further, even when all desirable information regarding outcome and probability is available, human decision makers do not follow the core axioms of EU (Tversky & Kahneman, 1986). Across many examples, Tversky & Kahneman, (1986) described the framing effect: how the context or presentation of a choice, without any manipulation of the outcomes or probabilities, can change preference. For example, when an option is framed in terms of losses it is less popular than when it is framed in terms of gains, despite no real change in probability or outcome, suggesting people are risk averse. This evidence is problematic for any theory which would describe humans as rational, by the standard of normative theories. It also diverges from the belief that rational choice must be a beneficial and a favourable trait in a competitive environment (Tversky & Kahneman, 2017). Nonetheless, there can be no doubt that most people are successful, competitive decision makers (Simon, 1990). What this evidence points towards is the need for a separate approach, or multiple approaches, to decision making which work well in daily life, but are exposed when asked to compete within the perfectly rational bounds of economic games (Gigerenzer & Gaissmaier, 2011). Arguably, if researchers only test theories based on the so-called norms of decision making then they can only ever show the human decision maker as less than or equal to them (Hammond et al., 1987). Prescriptive theories aim to predict how humans actually make decisions, without the need to directly contrast performance with normative predictions.

One common approach in prescriptive theory is to consider different strategies for making a decision. A decision maker could apply strategies appropriate for the kind of decision that they are making, based on the information available. Most of the decisions people make in their daily lives, whether trivial or of critical importance, are made under varying degrees of uncertainty. Further, real world decisions are often more complex; information needs to be actively accumulated and sought out among distractions. It may be the case that more than one approach is needed to guide us through this kind of environment. Epstein, Pacini, Denes-Raj and Heier (1996) consolidated some of the many approaches that have been suggested and reduced them to two: the intuitive-experiential and analytical-rational systems. In their current form, these are commonly referred to as System 1 and System 2, respectively (Evans, 2003; Stanovich & West, 2000). System 1 is a fast and intuitive amalgamation of sub-systems which collectively work as a domain general problem solver. System 2 is more deliberative, consciously available and abstract–it is capable of simulating events that have not been experienced before. A dual-systems approach may allow people to perform efficiently under varying degrees of uncertainty.

While the dual-systems approach may explain how human decision-making performance can vary so readily in different situations, it presents a new problem: how is it decided which system is used to make a decision. This can be labelled as the problem of strategy selection. Beach and Mitchell (1978) began their investigation of strategy selection before the dual-systems approach was as understood as it is today. Nonetheless, their Contingency Model for strategy selection can be applied to Systems 1 and 2. They suggested that individuals select strategies based on their characterisation of the decision task. These characteristics may include time and resource constraints, task importance, and familiarity. Based on these characteristics, the costs and benefits of applying each strategy can be assessed. From this, it can be predicted that variations in task characteristics will manipulate strategy selection. Christensen-Szalanski (1978) tested this prediction by varying the available time and importance of decision tasks. Of three strategies available (high, medium and low resource), participants chose to spend more resources (analogous to System 2) only when they were not confident enough in low resource (System 1) strategies. The conclusion then that task characteristics drive strategy selection is quite convincing (see also: Payne et al., 1988). However, a new problem is presented by any attempt to decide strategy based on cost/benefit analysis: how is this decision made? New lotteries are formed from the probabilities of strategy success and utility is measured against confidence in solutions. Rapidly, this problem takes the form of a decision made under uncertainty to be solved by SEU, which has already been dismissed. If SEU theory is not the solution here, then the problem becomes recursive.

One way of escaping recursion is to rely on a stable, external source, like the environment. This way of thinking has found success in other fields, such as visual perception and memory (O'Regan, 1992; Tulving & Schacter, 1990). In this sense, Beach and Mitchell (1978) were correct about task characteristics influencing decision strategy. In a similar way, Gibson's (1979) ecological approach to visual perception suggested that the environment possesses characteristics which describe what actions are available. These are called affordances of the environment, which need only be perceived by an agent for a course of action to be chosen. Based on this idea, an ecological model of strategy selection can be developed. Marewski and Schooler (2011) proposed that each decision strategy has a cognitive niche which defines the set of affordances for which it is appropriate. They suggest that, while cognitive niches can overlap, the problem is greatly simplified by reducing the number of potential decision strategies. This does not quite eliminate the problem of cost/benefit analysis-based strategy selection, but it presents it in a way that may be solved by cues from the environment.

The decision-making research described so far has been highly testable and theory driven. The axiomatic approach of classical decision making has allowed it to be built on over many decades. Unfortunately, in psychological research this scientific rigour comes at a cost. In this case, the prescriptive models, which attempt to explain real human behaviour, are difficult to apply to most situations; they explain performance in simple economic games and can be used analogously to evaluate daily decisions like “tea or coffee?”, but falsifiable tests of complex human behaviour are challenging (Roy, 1993). The trade-off between how complete the explanation and how much is explained is a common theme in psychological research and forms one of the major challenges of unified theory of behaviour. This subject will be revisited in more detail in my review of naturalistic stimuli in research.

2.2.2. Naturalistic Decision Making and expertise

Naturalistic Decision Making (NDM) is a scientific movement that approaches the investigation of decision making in a very different way to the classical research described in the previous section. The formation of NDM was a direct response to what researchers felt was a limitation in the applicability of classical decision making theory. One prominent example would be the development of ‘decision aids’ by medical practitioners (O’Connor et al., 1999). These decision aids give patients information about the different choices and the costs and benefits of each, with the aim of improving decision making under uncertainty. However, a meta-analysis and review of 105 studies on the efficacy of decision aids concluded that they do not change patient decision making (Stacey et al., 2017). Only mixed findings suggest that patients are more conservative about invasive surgical options but that effect was only observed after removing a major study with an opposite finding from the analysis (Schwartz et al., 2009). Another interpretation would be that when more information is had, risk aversion is more prominent.

That is not to say that they have no effect, as patients both feel, and are, more informed. Similar limitations of decision aids have been found experimentally (Yates et al., 2003) and are reported to be critical in the formation of the NDM field (Klein, 2008; Lipshitz et al., 2001). Notably, around the time the NDM movement began, other applied researchers were calling for a second scientific approach that dealt with the reality of decision making for the purposes of improving decisions, rather than contributing to the axiomatic field of classical decision making (Roy, 1990, 1993). Coincidentally, Roy suggested this be called the science of 'decision aid' (later becoming Operational Research and Decision Aid).

The cornerstone of the NDM movement was a study that aimed to investigate decisions made by experts under time pressure (Klein et al., 1986). Their investigation involved interviewing 26 highly experienced (mean average of 23 years' experience) firefighter fire ground commanders about non-routine situations they had been involved with in the last year. By the researchers' own report, the format and analysis of interviews was designed to investigate how experts decide among alternative courses of action. In this sense, the planned research was not dissimilar to applied classical decision making. The primary phenomenon of interest was decision points, defined as points in time in which alternative courses of action were available. One example given was the decision to evacuate a building or continue evaluating the source of smoke. The interviewed fire ground commander had to evaluate the risk of the fire spreading up from where it has started in the basement, versus the risk to occupants. During the interviews and in later analysis, the researchers realised that the fire ground commanders were not explaining their deliberation in a way that fit expectations. They found that reports of parallel evaluation of options rarely (12%) occurred. Rather, at 80% of decision points the fire ground commanders reported a more serial approach where only one option was considered at a time. To clarify, these decision points did not meet the original definition but were included anyway. Further analysis suggested that the commanders were recognising more standard scenarios in aspects of the current one and using those to generate the best course of action. This led to the development of the Recognition-Primed Decision (RPD) model (Klein, 1993). This idea has been described similarly elsewhere as the recognition heuristic (Gigerenzer & Goldstein, 1999, 2011)

There are several criticisms that can be made about Klein et al. (1986) which the authors acknowledged. Most notably, the use of interviews meant that bias and error was introduced by both the participants and the interviewer. The median time since the incidents discussed in interview had occurred was less than one year, but over a third had occurred more than three years prior. The authors claim that the critical incident report method used (Butterfield et al., 2005; Flanagan, 1954) is effective for non-standard events. However, in the time between incident and interview it seems likely that the events will have been written about and discussed many times, especially if it was non-standard, which means there will have been multiple opportunities for false memories to form (Strange & Takarangi, 2015). This is an accepted caveat of many studies using critical incident reports. A second criticism is that even with perfect recall or direct observation, the methods used in this study assume that deliberation is made available to consciousness (Nisbett & Wilson, 1977). Again, this is a common caveat in qualitative research, but does not necessarily prevent theories being generated which can later be tested (Ericsson & Simon, 1980; Kvale, 1994). Ultimately, the main outcome of the study was not their empirical findings and was not dependent on methodology (Klein et al., 1986). Rather, it was the suggestion that the classical model of decision-making being used was not suitable for their research and that therefore a new model was needed (Klein, 1993).

An illustration of the RPD model can be seen in Figure 1. The model describes how it is possible for experts to immediately come up with one course of action and carry it out without simultaneous evaluation of other options (Klein, 1993). Note, the RPD model was not designed to generalise to all decision making; it explains the performance of experts whose experience affords them the possibility of applying past solutions to the current situation. Even so, the RPD model's reliance on cues from the environment is quite like the ideas presented earlier, in the review of classical decision making. Rather than the environment guiding decision strategy selection, RPD would suggest a higher-level interaction with the environment. This is in keeping with the NDM movement's grounding in the ecological framework (Militello et al., 2017). For example, "It is not true that "the laboratory can never be like life." The laboratory *must* be like life!" (Gibson, 2014, p. XV). Alongside the recognition circuit, central to the model, there are two complementary components. They are complementary because they are only required when recognition fails. They also do not feature in the 'simple' version of the RPD model which is claimed to explain most decisions by experts (Klein, 1993).

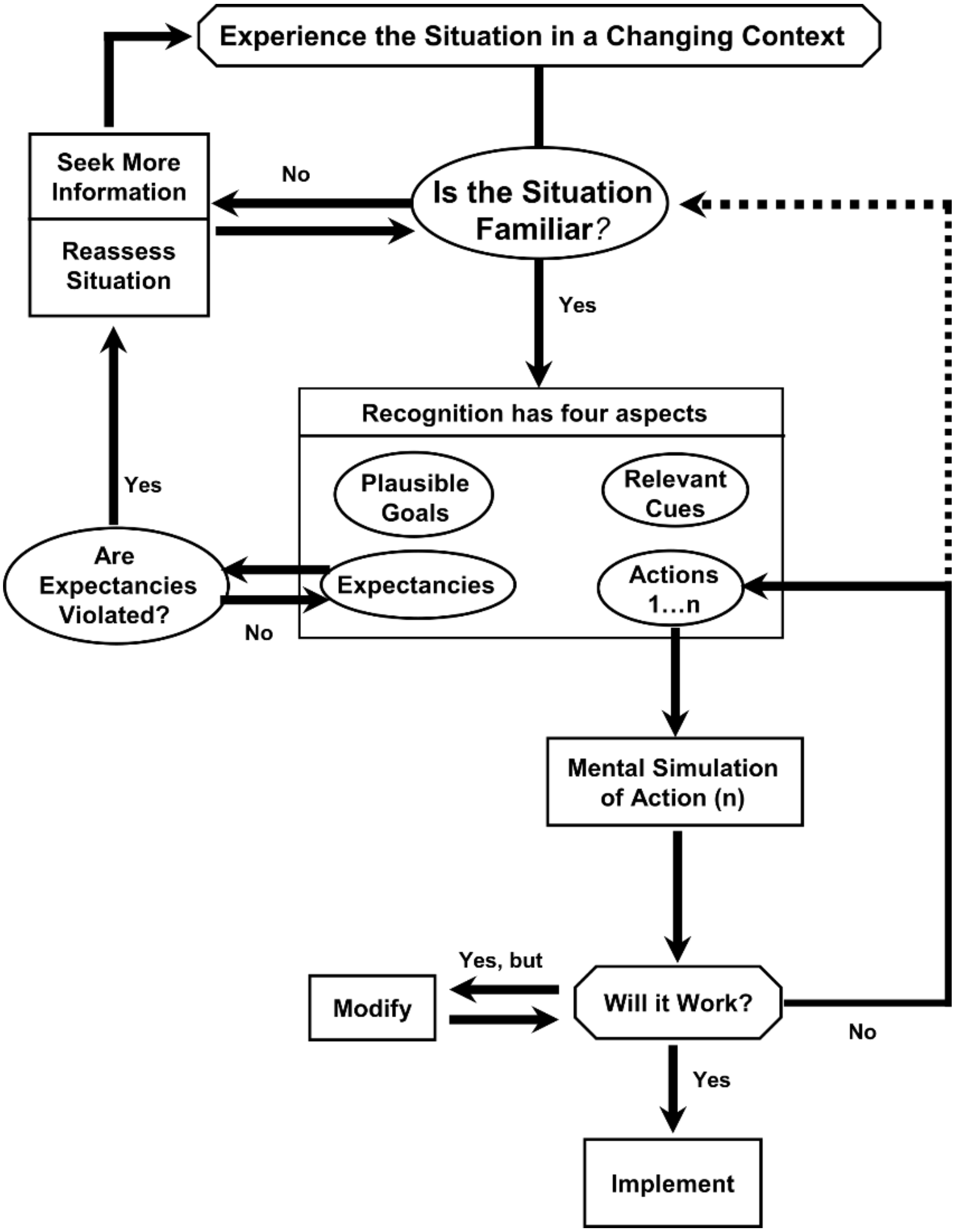


Figure 1. The Recognition-Primed Decision model. The model describes how experts make a decision using sequential evaluation, rather than simultaneous evaluation of options. This figure has been reproduced from Klein (2008). Copyright © 1993 by Ablex Publishing Corporation. Norwood, NJ.

One of these is the action queue, which can be seen in the bottom right circuit of Figure 1. The model suggests that once a situation has been recognised and a course of action decided, it is first simulated to see if it would work. If it does work, then it is acted on, but if not, then the next action is considered. Klein et al. (1986) added this into the model based on the verbal report of the fire ground commanders, using the example given earlier about choosing the right harness. The interviewee reported assessing each harness in turn: if it did not work for the current situation, they moved to consider the next available option. The authors interpreted this as a serial evaluation of the options discussed. This interpretation is open to criticism, as it is hard to imagine how the interviewee would have reported parallel evaluation of options. Moreover, their reports were based on introspection–the general limitations of which have already been mentioned–from which it may be difficult to experience anything other than sequential thought. Despite this, the authors identified the action queue as an important phenomenon and included it in their RPD model.

Another complementary part of the RPD model addresses what happens when the situation is not familiar (top left of Figure 1). The solution provided is that if the current situation is not familiar then more information is needed. In realistic, complex situations new information is constantly being made available and so it is feasible that delay may allow a decision maker to identify new patterns to guide them to an action. This is supported by the data collected from fire ground commanders, as they reported few split second decisions, although most were under one minute (Klein et al., 1986). What is not made clear in the model is what happens when someone lacks the experience to allow them to find a situation familiar. Admittedly, the aim of the RPD model is not to explain those situations. In my opinion, the strengths of the RPD model belong to the simple version that only includes successful recognition. Neither the action queue nor information seeking are critical but are needed to make the model more complete by covering failures of expertise.

A key characteristic of NDM studies is the use of expert participants. Evidence for the RPD, or very similar alternative models (e.g. RAWFS, the reduction, assumption-based reasoning, weighing pros and cons, suppression, and hedging heuristic; Lipshitz & Strauss, 1997) has come from studies on nurses (Currey & Botti, 2003), chess players (Johnson & Raab, 2003; Klein et al., 1995), members of the Israeli Defence Force (Lipshitz & Strauss, 1997), and police officers (Hine et al., 2018; Suss & Ward, 2018). Unsurprisingly, given the work presented in this thesis, we believe that the use of experts as participants is an excellent tool for investigating decision making. However, the difference in culture between researchers and professionals (e.g. police or the military) must be respected by both parties (Cockbain

& Knutsson, 2015). Many NDM publications target application in the military, including specially adapted versions of the RPD model (Klein, 2008; Ross et al., 2004). Because of this, it is worth noting that, while not declared as a conflict of interest in recent papers (e.g. Klein, 2015; Klein et al., 2018), the founder and a major proponent of NDM is associated with three training consultancy businesses in the military sector (Klein, 2015). This observation is not meant to detract from their work, or NDM as a whole, but must be considered when evaluating the effectiveness of training methods developed from it (e.g. Klein et al., 2018).

2.2.3. Neuroscience of decision making and the reward circuit

Research into the neural correlates of decision making has been guided by the behavioural investigations discussed above. In decision making neuroscience, experiments in the form of simple economic games or artificial, contrived experiments are typically used. To my knowledge no NDM studies have used neuroscientific methods. The topic of neural correlates of decision making is extremely broad and so I will limit this review to research related to neural correlates of human decision making under varying degrees of uncertainty.

The brain regions and networks associated with decision making under uncertainty are necessarily related in part to low level functions that are not unique to humans or complex decision making. Animal studies using invasive techniques, including lesion, electrical stimulation and pharmacological stimulation/inhibition (e.g. Blaha et al., 1996; Carr & White, 1983; Olds & Milner, 1954), have been used extensively to study decision making functions in the mammalian brain. The basal ganglia, which were historically associated only with motor and sensory-motor behaviour (Nauta & Mehler, 1966), have been found to contain multiple distinct circuits for different functions (Heimer et al., 1982; Mogenson et al., 1980). This distinction led to the discovery of a cortical-basal ganglia network associated with risk assessment and reward evaluation, called the reward circuit (Alexander & Crutcher, 1990; Haber et al., 1995; Joel & Weiner, 1994). Investigation of the human reward circuit *in vivo* has been motivated by these animal studies and enabled by advances in non-invasive functional neuroimaging techniques (Draganski et al., 2008). These investigations show that the core basal ganglia structures of the reward circuit are the ventral striatum and substantia nigra pars compacta. Their indirect projections to the prefrontal cortex result in feedback from the ventral-medial prefrontal cortex (vmPFC), orbital frontal cortex, and dorsal anterior cingulate cortex. For a review of this work see Haber and Knutson (2010). However, the reward circuit is interconnected, sending and receiving input from many other networks and structures, including the amygdala, hippocampus, dorsal prefrontal cortex and limbic-basal ganglia networks associated with motor behaviour. Disentangling these connections is challenging but supports the reward circuit's role in complex decision making; integration with these networks is essential for receiving information to guide decision making and, later, the enactment of the decision (Haber, 2017; Khalighinejad et al., 2020).

As discussed in the previous section, there is great variation both within- and between-individuals in decision making performance, but on the whole people are capable decision makers. The reward circuit and its connections have been targets of research into this variation in performance. The minimally invasive techniques used on humans mean that much of this research studies patients with localised damage to their brain. Bechara et al. (1994) developed the Iowa Gambling Task (IGT) to analyse decision making under varying degrees of uncertainty in patients who appeared to have deficits in decision making that was affecting their daily life, and which were believed to be caused by damage to their brain. In the IGT four decks of cards are presented, each with varying probabilities of giving rewarding or punishing cards when chosen. Because of this, some decks are better or worse than others, but this information is not explicitly available to participants. Instead, decks must be sampled and compared by reward. A "win-stay/lose-shift" strategy best predicts control group performance, suggesting they rely on implicit processing of feedback from earlier actions (Damasio et al., 1996; Worthy et al., 2013). However, patients with lesions within the vmPFC do not successfully implement that strategy. Bechara et al. (1994) found that the patients sought immediate reward at the cost of delayed punishment, causing them to underperform. Interestingly, in tasks where immediate reward seeking is beneficial, patients with vmPFC lesions outperform healthy controls who prioritise risk aversion the most (Shiv et al., 2005). Another study using this patient group compared the IGT performance of patients with vmPFC lesions against patients with bilateral amygdala damage (Bechara et al., 1999; see also: Brand et al., 2007; Weller et al., 2007). While both groups underperformed, reactionary emotion responses, measured by skin conductance, were observed in the vmPFC lesion group, but not the amygdala damage group. The authors of this research suggested that this was because the amygdala is involved in generating the emotion response to a rewarding/punishing card (for an alternative perspective see Pessoa, 2011), but the vmPFC is necessary for secondary emotional processing and for actually applying this to future decisions. Together, these findings support the Somatic Marker Hypothesis (SMH) which states that when acting under uncertainty, somatic markers direct behaviour (Damasio et al., 1996). Here, somatic markers are defined as emotional states assigned to an event, either in immediate response (primary) or as recall (secondary). System 1 is often discussed in terms of a 'gut feeling', emotional response which causes people to act one way or another (Zeelenberg et al., 2008). For this reason, it is unsurprising that regions associated with emotion processing have been related to decision making, as in the SMH. Also unsurprising is that the human stress response affects some of these same areas.

2.2.4. Stress and decision making

Terms such as 'anxiety' or 'pressure' are variably used to describe the same phenomenon: stress. Stress is a natural, generalised response to the perception of some aversive stimuli. Selye (1936) described this response as the General Alarm Reaction after observing the physiological response of rats to nocuous, or harmful, agents. He proposed that this response was an attempt by the rats to adapt to the new challenge they were presented with. The use of the term 'general' was core to this description because it contentiously suggested the same response applied to all kinds of aversive stimuli, including psychological (Lazarus, 1993). After three decades of research and debate, it was commonly agreed that the General Alarm Reaction, or stress response, applied to psychological stressors as well (Mason, 1968b, 1968a). While this is still widely accepted, a more recent review of the neural correlates of psychological stress shows that this is not conclusive (Dedovic et al., 2009). Further, Brisinda et al. (2015) demonstrated it was possible to differentiate physical and psychological stressors using quantitative analysis of heart rate variability. By coincidence with the broader topic of this review, they used electrocardiogram data from police officers while they conducted various tasks, including firearms and tactical training, exercise, and normal daily activity. Building on early work, stress has also been redefined as having three stages, the stimulus, perception of the stimulus, and physiological response, all of which need to be present and which can have variable intersection with other cognitive functions (Koolhaas et al., 2011). This is important, because the physiological stress response pathways are not exclusive to stress.

The human stress response is dominated by two main pathways: the sympathetic adrenomedullary system (SAM system) and the hypothalamic-pituitary-adrenal axis (HPA-axis). The SAM system is a rapid neural pathway whereas the HPA-axis is a slower, hormonal response. Following the first two stages of stress, the stimulus and perception of the stimulus, the hypothalamic paraventricular nucleus increases secretion of corticotrophin-releasing factor (CRF) and vasopressin. Both CRF and vasopressin are transported to the anterior pituitary gland via the hypophyseal portal system where they both promote the release of adrenocorticotropic hormone (ACTH) which is then circulated. ACTH binds to available receptors on the adrenal glands and promotes the release of glucocorticoids into the bloodstream, the most prominent of which is cortisol. Cortisol has wide reaching effects and can bind to many cells within the human body; it causes gluconeogenesis (the biological pathway to increase blood sugar levels) and reduces immune system function. The SAM system also originates in the hypothalamus, but the response happens within the autonomic nervous system. Specifically, sympathetic preganglionic neurons stimulate the adrenal medulla, causing it to release adrenaline and noradrenaline hormones. With immediate effect, these increase heart rate, blood pressure, and stimulate gluconeogenesis. Overall, the effect of these two systems is to adaptively free up resources to tackle the perceived stressor. There are maladaptive effects as well, especially if the stressor is persistent, as homeostasis is widely disrupted.

The physiological stress response also has direct effects on the central nervous system. One part of this effect is for regulation, to ensure that hormone levels return to normal at the cessation of the stressor. Cortisol levels remain elevated for up to one hour and peak between 21-40 minutes (Dickerson & Kemeny, 2004). This has been shown to be effective in studies where a stress inducing task is used before the experimental session, rather than the task itself being stressful (Porcelli, 2014; Smeets et al., 2012). However, regions of the brain associated with emotional processing, including the hippocampus and amygdala in the limbic system, and the prefrontal cortex, have receptors for glucocorticoids which promote metabolic changes within them (Pruessner et al., 2010). Specifically, increased cortisol binding reduces activity in the hypothalamus, hippocampus, amygdala and vmPFC (Lucassen et al., 2014). Increased cortisol in response to stressors has been found to increase dopamine release in the ventral striatum and vmPFC within the reward circuit (Joseph et al., 2003; Pizzagalli, 2014). Behaviourally, this manifests as increased hedonic motivation to resolve the stressor (Ironside et al., 2018). Interestingly, anhedonia may occur as a response to chronic stress, as in the case of major depressive disorder and post-traumatic stress disorder, due to compensatory change in activation in the related brain regions (Enman et al., 2015; Hollon et al., 2015). The effects of physiological stress on decision making, via the biological mechanisms described, support further research into the effects of acute stress on decision making under uncertainty.

It is important to understand how increased motivation in response to acute stress manifests as change in behaviour and subsequent performance. It is commonly reported that acute stress leads to increased risk taking and reward seeking (Starcke & Brand, 2012, 2016). This has been found in research which has compared the performance of stressed participants with controls when completing various decision making tasks. Lighthall et al. (2009) reported that stressed men performed best at the Balloon Analogue Risk Task (BART). The BART rewards participants for continuing to blow up a virtual balloon but punishes them if the balloon pops. Because of this, a reserved strategy could never achieve the highest score. If the task were adapted to punish risk, say by increasing the variability of the point when the balloon pops, then stressed male participants would likely perform worse. This study also found that women were risk averse when stressed, as shown by their poorer performance on the BART (see also: Lighthall et al., 2012; van den Bos et al., 2009). Similar effects have been observed in studies of police recruits (van den Bos et al., 2014), although they did not use tasks directly related to policing. However, the finding is not consistent and may be related to sensitivity to stress inducing tasks, oral contraceptives, or more general differences in decision making unrelated to stress (Starcke & Brand, 2016). Even in decision making tasks with less uncertainty, such as the Game of Dice Task (GDT; Brand et al., 2005), stress has been found to increase risk taking. In the GDT participants must gamble on between one and four dice rolls appearing in a set of eighteen simulated rolls. They are rewarded if their numbers appear, depending on how many dice they chose to predict. Starcke et al. (2008) found that stressed participants performed badly at this task and that performance was negatively correlated with cortisol levels. Similar findings have been found for other tasks, such as the Iowa Gambling Task (Preston et al., 2007) and Cambridge Gambling Task (Porcelli & Delgado, 2009). This trend of increased risk and reward seeking behaviour makes sense when considering the purpose of a stress response: to help overcome a new challenging situation. If this situation could be overcome with normal, conservative strategies then no change would be needed.

2.3. Police performance

The unique position of police officers in society means that they are of potential interest to researchers working in many different fields. Given this interest there are surprisingly few studies of police officer behaviour. There are several possible reasons for this deficit. The first was briefly mentioned in the introduction to this thesis and is the challenge of meeting methodological standards while maintaining the ecological validity required to investigate a group who only exist outside the laboratory (Hope, 2016). Suss and Boulton (2019) summarise some of the other, more logistical, challenges in a recent review. They suggest that the cultural divide between research and law enforcement can be challenging to overcome. Simply, police are far more exposed than researchers and so understandably must see some benefit to participating in research when they take on most of the risk. Researchers also take on some risk, as their research may be politicised or misinterpreted (Hope, 2016). Other logistical challenges include access to participants and funding (Angel et al., 2012). Of course, these problems are not unique to law enforcement research and apply to many research areas working with specialised groups. Note, police officers are not a small or rare group–in England and Wales there are ~200,000 police officers (Home Office, 2019b) –but their work is demanding, time-constrained and mobile. Support from leadership and administration is therefore essential for participant recruitment. All this is not to say that there is no research at all, but to provide context for some of the limitations in theory and methodology of the work discussed in this section.

When discussing police performance, the first thing to do is reduce what is a multifaceted job into areas that distinguish it from others (Suss & Boulton, 2019). Police officers are involved not just in the protection and prevention of crime, but the legal processes surrounding those actions. For this reason, police have been studied as part of investigations into eye witness testimony (K. A. Wade et al., 2018), officer conferencing (Hope et al., 2013) and interview techniques (Gabbert et al., 2009). These investigations share a common motivation with our own: to better understand the behaviour of individuals who have additional powers in society and to contribute to their effectiveness. However, the research methods and cognitive processes studied are not directly related to aspects of performance we are interested in and are out of the scope of this review. Our interests relate to another salient aspect of policing: the use of firearms. The use of firearms varies greatly between police forces, as does the training provided–as little as 60 hours in the USA (Reaves, 2009, 2016) –but they undoubtedly form an important part of policing. Reports of shooting performance are also understandably varied. Police equipped with firearms usually must pass a minimum standard of static shooting performance which is commonly achievable (e.g. 70% minimum hit rate with 99.7% pass rate in Illinois, USA; Charles & Copay, 2003). Similarly, high static shooting performance (95%) has been reported in the Netherlands (Nieuwenhuys, 2011 in Nieuwenhuys et al., 2017). For real world police shooting performance, the largest dataset comes from the USA. A review of reports from many hundreds of incidents between 1980 and 1992 estimated between 27% and 60% hit rate across multiple police forces (Morrison & Vila, 1998). To my knowledge, no similar report has been made with respect to UK police forces. However, in England and Wales between 2010 and 2019, there were 61 incidents in which police firearms were discharged. Of these, 22 resulted in a fatal shooting (Home Office, 2019a; Inquest, 2020), so a similar percentage can be assumed.

While these statistics are important for informing policy and training, psychologists are more interested in the factors which affect shooting performance. Several studies supported by police in the Netherlands have investigated these factors. One study investigated how shooting under pressure affected shooting performance, with an aim to reduce the discrepancy between practice and real life shooting (Oudejans, 2008). They developed two shooting tasks which could be manipulated to be high- (against an opponent) or low-pressure (cardboard target). They first tested the performance of 17 police officers at both tasks and recorded their performance and heart rate. Heart rate was higher and shooting performance worse in the high-pressure version of the tasks. The participants then went through training, but half continued training against an opponent, while the others did not. After training, all participants completed the two tests again. They found that participants who trained with an opponent performed at the same high level in both low- and high-pressure tasks, while the other group continued to perform worse in the high-pressure task. While presumably expertise at shooting without an opponent is a prerequisite of shooting against one, this study highlights the importance of additional training in more realistic settings. A second study by the same research group used different shot execution strategies (step-fire or fire-step) to test whether inhibition of preferred response impacted anxiety and performance during high-pressure shooting (Nieuwenhuys et al., 2017). They found that when the less-preferred fire-step strategy was used, anxiety was higher, shooting accuracy was poorer, and aiming time was shorter. Taken with the results of their earlier study, their findings support the idea that anxiety has a negative effect on shooting performance. In further analysis of their data, the same research group has shown that individual differences on self-control measures predict shooting performance under pressure (Landman et al., 2016a). They suggest that self-control, measured as part of the Action Control Scale (Kuhl, 1992), was crucial, as differences in heart rate and reported anxiety could not explain the differences among participants. These studies are also good examples of some of the challenges to research on police performance and how they can be overcome.

The vast majority of incidents in the UK with a firearms response do not result in police discharging their firearms (0.064% in 2019; Home Office, 2019a). Arguably this is an indication that measurement of shooting performance does not fully represent police performance with firearms. Further, empirical data across multiple simulators suggests that there is no relationship between marksmanship and the number of decision errors made during the simulations (Blacker et al., 2020). Therefore, it is important that we understand the factors that contribute to the decision to shoot, whether in error or as an appropriate response to threat. One study compared participant performance at a computer game (Wii console) where the challenge was to shoot virtual perpetrators while avoiding virtual civilians (Biggs et al., 2015). They found that those with lower inhibitory control performed worse and were more likely to shoot civilians. Subsequent response-inhibition training (30 minutes go/no-go and 30 minutes stop-signal reaction time task) improved performance at the task relative to a control training task, supporting evidence that cognitive training can improve inhibitory control (Spierer et al., 2013). Note, that while the simulator was not realistic, measures extracted from the game about marksmanship and shooting performance are highly correlated with those from a more realistic military training simulator (Blacker et al., 2020). Preliminary findings from the study show response inhibition training (forms of which do exist in police firearms training) could lead to fewer civilian casualties. Their findings have been replicated in a study using police officers as participants in a more realistic task (Hamilton et al., 2019), but the long term effects of cognitive training are unclear. Another study measured the effect of fatigue on performance of military infantry in a variety of tasks, including realistic shooting exercises (Nibbeling et al., 2014). Despite collecting data from military infantry, the task was closer to that of police training. One of the tasks involved deciding whether to shoot an actor who surrenders or attacks. They found that participants made more anxious by exercise-induced fatigue and the possibility of being shot back at with training ammunition (intense pain) were more likely to shoot a surrendering actor. This effect was only borderline significant but given their reported large effect size (f = 0.48) with a small sample (N = 22) it is noteworthy. Further, they confirmed their anxiety manipulations using combined physiological and psychological measures. These studies further support the impact of anxiety on police performance and extend the effect to the decision to shoot.

The context of a decision to shoot or not must also be considered. There are several ways in which an AFO can be deployed in the UK. They can be deployed by a firearms commander in response to an incident or for a planned operation, or they can self-deploy if they encounter a situation that demands it (College of Policing, 2013). These circumstances can reasonably be extended to other police forces. The 'shoot/don't-shoot' studies discussed so far largely apply to self-deployment situations as participants do not rely on intelligence from a control room. Note, AFOs must always self-authorise if they decide to discharge their firearm, so the 'shoot/don't-shoot' studies do still apply. Taylor (2019) addressed the question of whether intelligence provided from the control room influences police officers' decision making–they call this 'dispatch priming'. Specifically, they tested the hypotheses that incorrect information about what a perpetrator is holding (e.g. a firearm) would increase shooting error and correct information would decrease shooting error. To do this, they first gave participants information about the scenario they would be completing. Briefly, they were going to investigate a possible trespassing in an empty home. They were then either given an update to say the perpetrator had been see "holding a handgun", an update saying they were "using their phone", or, in an ambiguous condition, they were given no update. When they were completing the scenario, the perpetrator reached into their pocket and rapidly pointed their phone or a handgun at the participant, depending on the experimental condition. The only measure taken was whether participants decided to shoot their training pistol at the perpetrator or not. Each participant completed only one scenario with a randomly assigned update and scenario outcome, but they collected data from 306 police officers. They found that regardless of the update, all participants who were assigned to the handgun scenario shot the perpetrator, which suggests either the task was not sensitive to influence from the control room, or a reliance on direct information. Unfortunately, no reaction times were reported, which may have varied between update conditions. However, findings from the phone scenario are compelling. Under ambiguous conditions, they found 28% of officers shot the perpetrator, increasing to 62% when given incorrect information that they had a handgun, and decreasing to 6% when correct information was given. These data support their hypotheses about dispatch priming and the relevance of prior information to responding to later stimuli. Stress may inhibit top-down, goal-directed processes (Eysenck et al., 2007), so it is possible the dispatch effect would be reduced in real life situations, for better or worse. This is supported by a study conducted on 80 police officers in the USA which looked at the effects of stress (induced by an adapted version of the Trier Social Stress Test; Allen et al., 2017) on error rates in a 'shoot/don't-shoot' task (Akinola & Mendes, 2012). The task had two variables: the perpetrator could be black or white and they were either armed or unarmed. The average error rate across conditions showed a bias towards not shooting armed white perpetrators, suggesting a prior bias was affecting their decision. They found that participants who had a greater increase in stress relative to baseline made fewer errors overall, and the relationship was strongest for trials with a black perpetrator. The authors' interpretation is that this indicates greater vigilance to threat when stressed, citing reports of association between fear and black people in the USA (Cunningham et al., 2004; Nosek et al., 2009). However, considering more recent findings, their data are better explained by greater reliance on bottom-up processing of threatening stimuli and suppression of bias.

Stress is clearly important to police performance and decision making. In addition to the research mentioned above, short-term occupational stress may affect cognitive performance (Giessing et al., 2019; Gutshall et al., 2017), cause avoidance behaviour, and negatively impact arrest and self-defence skill (Renden et al., 2014, 2017), and increase aggression (Queiros et al., 2015). Because of these effects and a general interest in the wellbeing of police officers, recommendations have been made for how stress can be reduced using effective training (Ekman, 2015; Papazoglou & Andersen, 2014). Another approach, which many have called for, is to use more realistic training to help prepare police officers for how their behaviour might change when under stress and to help inform selection procedures (e.g. Connelly et al., 2019; Giessing et al., 2019). It will be challenging to reproduce the levels of stress experienced in a real-life firearms incident, but the current trajectory of improvements in task realism (aided by increased accessibility and quality of virtual reality technology) will help researchers work towards that aim. Already, training scenarios used by police have been shown to produce comparable physiological responses to real world observations (Armstrong et al., 2014). When interpreting any of the results presented that depend on task-induced stress, serious consideration must be given to possible non-linear effects of stress, described by the Yerkes-Dodson law (Yerkes & Dodson, 1908). Research in this area is not limited to police decision making, and the wider literature on stress and decision making must be considered alongside it.

2.4. Naturalistic stimuli

The conflict between experimental methods which maximise internal validity versus those that maximise ecological validity is not unique to decision making research. However, classical and naturalistic decision making do represent the extreme ends of the scale in psychological research. Because of this, any positive evaluation of one becomes a negative of the other, and vice versa, resulting in a zero-sum trade-off (Figure 2; Loomis et al., 1999). In brief, laboratory-based investigations can be criticised for their poor ecological validity because of the many differences that lay between the tasks and situations that participants of these experiments encounter, and the everyday lives of people. Scepticism regarding ecological validity does not mean that the findings of research in the lab diverge from what would be found in the real world, only that they are untested. Likewise, experiments or observations that lack control of variables or use fewer objective measures can only make limited claims about what they measured. Again, it is possible that hypotheses generated from that type of research can be verified in the lab, but it is not simple to do. For these reasons, it is important that techniques are developed which can be used to investigate human behaviour as naturally as possible, while maintaining experimental control. From this idea, researchers are beginning to use more naturalistic stimuli in their experiments.

A naturalistic stimulus is one that the experimenter has complete control over, but which through the use of technology, or careful design, is perceived by participants as the same, or similar, to something they might perceive in the real world. For example, a photograph of a scene or a recording of speech would be naturalistic. The opposite would be the more common artificial stimulus, such as a Gabor patch (Fredericksen et al., 1997) or the beep of an auditory oddball task (Squires et al., 1975). One stimulus can also be considered more natural or artificial than another. While it is generally intuitive to what category a stimulus belongs, it is important to characterise what features make a stimulus more or less natural. Ideally, these characteristics should generalise across sensory modalities.

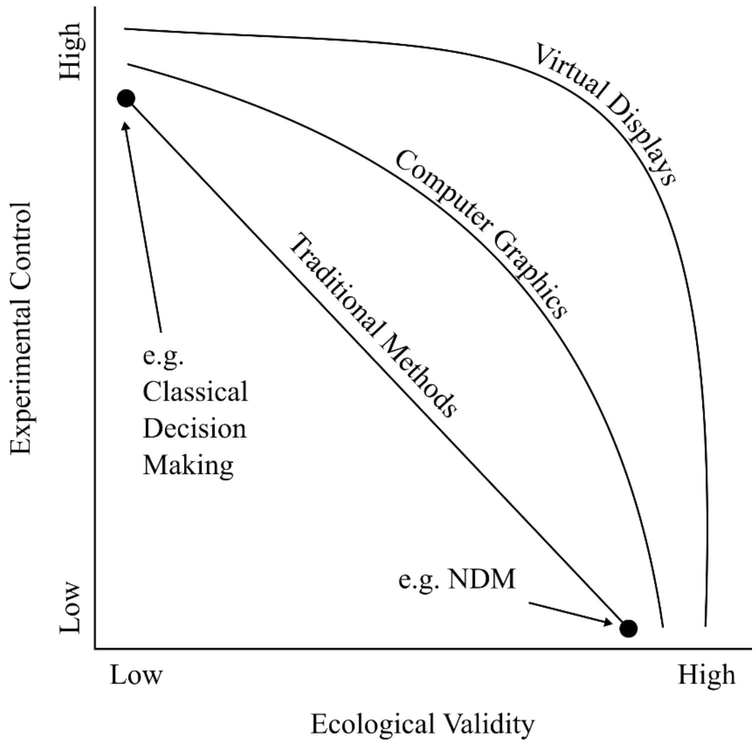


Figure 2. Zero-sum and non-zero-sum trade-offs between ecological validity and experimental control. The benefits of using virtual reality in research can be seen by the reduced trade-off between experimental control and ecological validity. This figure has been reproduced from Loomis et al. (1999) and amended with relevant examples. Copyright © 1999 by Psychonomic Society, Inc, Chicago, IL, USA.

One such characteristic is whether the stimulus is continuous or discrete. We may perceive speech and photos as discrete words and objects, respectively (Lieberman et al., 2005; Wagemans et al., 2012), but speech rarely has breaks between words and photographs do not have explicit boundaries between objects–they are both continuous. In contrast, artificial stimuli are more discrete. This is emphasised by the way they are presented in experiments, where each stimulus might be flashed up in turn against a mid-grey background. From this idea, Sonkusare et al. (2019) observed that naturalistic and artificial stimuli can be characterised by the shape of power spectral density (PSD) functions in the relevant frequency space. For example, consider an experiment where the stimuli are checkerboard patterns with check size of ten minutes of an arc (mins; 1/60th of a visual degree) flashed four times a second (as in Slooter & Norren, 1980). The power spectral density would show a sharp peak at 0.25Hz for temporal frequency and at six cycles per degree for spatial frequency. Observations of nature do not tend to have these sharp peaks, in favour of a smoother, ‘1/f-like’ power spectra, also known as ‘pink noise’ and ‘flicker noise’ (Johnson, 1925; West & Shlesinger, 1990). One common example given of pink noise in nature is EEG. However, I would argue that the signal of EEG is represented by divergence from pink noise. Measurements of current from an electrode on a semiconductor (like the scalp) will generally have pink noise (Caloyannides, 1974) which is not dependent on the signal (although see Bak et al., 1987). Regardless, the pattern is seen so commonly in nature that Sonkusare et al. (2019) argue that if stimuli used in an experiment are naturalistic then they too should have ‘1/f-like’ power spectra. Similar, although less explicit, observations have been made elsewhere by researchers using naturalistic stimuli for visual search (Wolfe, 1994) and social cognition (Redcay & Moraczewski, 2019).

There is evidence that the 1/f-like power spectrum is related to features of naturalistic stimuli that affect how they are perceived. Discussion of this evidence requires a brief summary of how power spectra are calculated. Any signal can be decomposed into frequency space as a combination of phase and magnitude across frequency, using the Fourier transform. Crucially, this can be done with no information loss; the signal can be reconstructed from the magnitude and phase components. Power spectral density is calculated from the square of the magnitude component (power) normalised to the width of the frequency bin (e.g. 1Hz). By only looking at the power spectral density of the signal from naturalistic stimuli, you cannot observe differences in the amount of information available at higher frequencies. However, we know that higher frequencies are essential for interpreting naturalistic stimuli. Oppenheim and Lim (1981) demonstrated this by reconstructing naturalistic stimuli, such as speech and images, using only the phase component on their frequency space decomposition. Phase-only reconstructions were made using the phase component and a uniform magnitude. They clearly demonstrated that phase-only reconstructions were highly identifiable. This is partially explained by the importance of high frequencies for defining edges. An edge can be conceptualised as a sharp boundary between two different values. Without high frequencies the boundary becomes smoother and therefore less distinguishable from other smooth changes. As an example, when magnitude at higher frequencies is reduced for an image it becomes blurred and appears less natural (Webster et al., 2002). Interestingly, blur also reduces phase coherence around edges, again suggesting the importance of phase for perceiving boundaries in continuous, naturalistic stimuli (Wang et al., 2004).

It is important that we understand characteristics of naturalistic stimuli because sensory and perceptual systems evolved in their presence (Felsen & Dan, 2005; Sonkusare et al., 2019). This means that we may be adapted to their specific characteristics (e.g. Schultz & Pilz, 2009). There are certainly more similarities than differences between naturalistic and artificial stimuli, which is why we have been able to learn so much about how we perceived the world using artificial stimuli. Rust and Movshon (2005) argue the case for artificial stimuli that their careful use has allowed us to understand a great deal about functions of the visual system that evolved for naturalistic images. In particular, they praise the efforts to understand the early visual system and improvements on the old standard model of vision, which was done using entirely simple, synthetic stimuli (Lennie & Movshon, 2005). They also highlight the importance of understanding the statistics of naturalistic stimuli if they are to be properly tested and warn against interpretation of null findings when fitting data from naturalistic stimuli to models built using artificial stimuli. Research from the same group shows how understanding of naturalistic stimuli can lead to new research questions, such as how continuous presentation of objects from multiple angles can be processed by the visual system and aid recognition (DiCarlo et al., 2012). This shows naturalistic stimuli can still be used in highly controlled experiments. In vision neuroscience, naturalistic stimuli have been used to investigate complex cells of the visual system, finding cells that respond selectively to congruent phase structures of natural images, rather than associated changes in magnitude (Felsen et al., 2005). Naturalistic stimuli have also been used in haptics research (Klatzky et al., 1985), olfaction research (Wachowiak, 2011), and translational research, using footage recorded from a cat's perspective (Kayser et al., 2003; Schäfer et al., 1992).

2.5. Virtual reality

2.5.1. Introduction to virtual reality

The term 'virtual reality' was coined in the late 1980s by Jaron Lanier as a catch-all phrase for virtual environments and the equipment used to present them. I have included a quote from Lanier talking about virtual reality at a computer graphics conference in 1989 below:

> *"It's a simulation of a reality that can surround a person that's created with computerized clothing. It's rather like the physical world in that it's an externally perceived reality that you perceive through your sense organs and the physical world… There is a special pair of eyeglasses called a head mounted display that you wear over your eyes and through the head mounted display you see pictures that are three dimensional stereo pictures… Then there is the glove.*

You put on a glove and the computer can measure your hand. You hold up your hand in front of your face and you see a hand. It lets you pick things up." (Conn et al., 1989, p. 1)

Lanier described a collection of technologies: head-mounted stereo displays; motion capture gloves; three-dimensional (3D) audio; and real time graphics. From a technological point of view, this definition of virtual reality still largely applies to the modern concept. However, the term is used inconsistently, both in the world of consumer technology and within academic literature. For example:

"Virtual reality makes use of virtual environments to present digitally recreated real world activities to participants via immersive (head-mounted displays) and non-immersive (2D computer screens) mediums." (Parsons, 2015, p. 2)

"...we use the term VR to mean 'a computer-generated world' and not just 'things viewed in a head-mounted display'" (Pan & Hamilton, 2018, p. 395)

"Experiences in virtual reality (VR) can be powerful–the user can feel as if he or she were actually ''present'' in the VR world." (Rosenberg et al., 2013, p. 1)

While there is overlap in these definitions, they vary enough that without supplementary knowledge they could be interpreted in different ways. This is a problem because it limits development of virtual reality as a research method. For example, a researcher may claim to use virtual reality in an investigation of social cognition and report that it was beneficial compared to other methods. Someone else may then apply this to their own work using a different definition of virtual reality and struggle to replicate the finding.

Fortunately, the problem of inconsistent use of the term virtual reality was recognised and arguably resolved in an early and widely cited article on virtual reality research (Steuer, 1992). They argued that definitions which take a technology-centric approach may work well enough for developers of that technology but are less useful in other areas. Of more interest to those studying virtual reality and using it as a tool for other research, is the experience of the user. This is important for a definition to stand the test of time, as technology changes but the intended experience does not. For example, television was not defined by the cathode ray tube, but by remote viewing. Along similar lines, Steuer defined virtual reality as an environment in which the user experiences 'telepresence', regardless of the technology used. Here, telepresence refers to the feeling of presence in an alternate reality which is remotely presented to the senses via some medium. While this medium is likely to be digital, as is the case for most, if not all, virtual reality equipment available today, it could plausibly be analogue (e.g. remote manipulators) or even biological.

Over time, the term telepresence has been developed and expanded on. In particular, within the context of virtual reality, the term telepresence is now often referred to simply as presence (Berkman & Akan, 2019). While others disagree with this simplification (Lombard & Jones, 2015), it is a reasonable assumption that discussions of presence in virtual reality is presented remotely through some medium. For clarity, I have kept the two terms distinct in this review. Further, presence has been divided into subsections, including social and co-presence (Slater et al., 2000). These refer to instances in which users feel that the environment is populated by other sentient beings (D. Roberts et al., 2006).

Steuer (1992) also addressed some of the terms often used in association with virtual reality, such as immersion and realism and reduced them into a hierarchy (Figure 3). Briefly, this hierarchy included vividness as a combination of the breadth (number of simultaneous modalities) and depth (resolution of modalities) of sensory input. Interactivity refers to the extent to which the user can modify the virtual reality in real time. Ideally, these would be equal to interactions available in the real world. Technology can facilitate this by reducing the latency of action effect, increasing the range of possible actions and allowing for actions to be mapped onto similar actions in the real world. Examples for each can be found in the original article. Additional variables have been suggested elsewhere (e.g. Slater & Wilbur, 1997), but are broadly covered under vividness and interactivity. Having these subcomponents allows for easier comparison between different types of virtual reality and study of their effect of telepresence. They also explain why people use the term virtual reality to mean many different things that vary in how vivid or interactive they are.

Steuer's (1992) definition of telepresence can be used to evaluate an experience. First, an objective prediction of telepresence based on measures of vividness, and interactivity. At a high level, these measures are simple to compare. For instance, if a display has a greater field of view then it will be more vivid and this should have a positive effect on telepresence (Slater, 1999). The second, and crucial part of the definition, depends entirely on the user’s subjective experience of telepresence. In order to validate the use of virtual reality, the first challenge is being able to measure presence. In doing so it is important to consider that presence is not unique to virtual reality. Any measure of presence in virtual reality can only be compared to presence in the physical world (Usoh et al., 2000). A perfectly vivid and interactive virtual reality should have the same level of presence as reality (perhaps greater, if science fiction such as Neuromancer [Gibson, 1984] can be considered), although the degree of presence can vary even then (Fontaine, 1992).

Figure 3. Technological variables influencing telepresence. This division of telepresence into subcomponents allows for easier discussion and evaluation of virtual reality experiences. Figure reproduced from Steuer (1992). Copyright © 1992 by Journal of Communication.

Interestingly, the relationship of presence with vividness and interactivity is not linear. It is possible that someone could feel more present in a simple, one-button game on their mobile phone than using a combination of the best virtual reality technology available today. On one side, this can be explained by bias to anthropomorphise even simple stimuli (Heider & Simmel, 1944). People may be more willing to accept agency based on competency than appearance (Koda & Maes, 1996). At the other side of the spectrum, is the Uncanny Valley (Mori [1970] in Mori et al. [2012]; see Figure 4). The Uncanny Valley analogy originally described the sharp fall in affinity for humanlike robots as their likeness approached reality and subsequent rise as that goal is achieved. Since then the analogy has been used more generally to describe the relationship of affinity and realism across multiple modalities (Pollick, 2010). Most commonly these are related to virtual humans or robots, but the Uncanny Valley may also exist for haptic feedback (Berger et al., 2018), audio (Grimshaw, 2009), and olfaction (Li & Bailenson, 2018). Another measure more unique to virtual reality relates to the ‘tele’ of telepresence.

For example, indirect measures of presence could be how dominant the virtual reality is over physical reality; how much the user feels they have visited somewhere else after the experience (Slater & Wilbur, 1997). Measurements of user's experience of 're-entry' from virtual reality into the physical world could also be considered (Behr et al., 2005). If they need to readapt their behaviour then this may indicate a high degree of presence in the virtual reality. Overall, these measures have demonstrated that subjective feelings of telepresence are related to the technology used (both hardware and software). Our understanding of this relationship is not complete but is in fact reassuring for those carrying out research using virtual reality, as simple applications may be as effective at promoting telepresence as those that approach reality.

Figure 4. The Uncanny Valley. Here, the analogy describes the sharp fall in affinity for virtual humans as their likeness approaches reality. All peaks would be exaggerated for moving stimuli (Gee et al., 2005). Figure reproduced from Mori et al. (2012). Copyright © 2020 by IEEE, New York, NY, USA.

It is common for research using virtual reality methods to measure the presence of participants in the virtual reality. This is commonly done by self-report, and several instruments have been developed for this purpose. For a detailed review of these methods, see work by Vasconcelos-Raposo et al. (2016). The first to be developed was the Slater-Usoh-Steed Presence Questionnaire (SUS; Slater et al., 1994).

It was developed based on the assumption that objective measures of vividness and interactivity should be positively related to a subjective measure of presence. This idea is supported by the relationship between ‘cybersickness’ and presence. Cybersickness is another term debated by researchers and may be indistinct from motion sickness and simulator sickness. Certainly, sickness can be caused by poor interactivity or vividness of a virtual reality. For example, if the breadth of the virtual reality only includes vision then mismatch between that and other senses which only receive input from the physical world may cause sickness. Therefore, it is unsurprising that cybersickness is negatively related to measures of presence (Weech et al., 2019), as cybersickness is itself negatively related to technological factors of telepresence (Kourtesis et al., 2019). A further issue with questionnaires is that they may bring about the state they aimed to measure (Slater, 2004). In this case, virtual reality users may not have considered the presence of their experience until questioned. Nonetheless, questionnaires about participants’ experiences may be useful as manipulation checks, especially when presence is manipulated.

Attempts have also been made to measure presence objectively. In two studies researchers have observed neural correlates of presence by comparing high- and low-presence experiences (Baumgartner et al., 2006, 2008). These experiences were two versions of a rollercoaster ride where presence was manipulated by increasing the amount of ascent and descent of the ride. Unfortunately, it is unclear which would have greater presence; by the standards I have described so far there is no clear difference in vividness or interactivity between the experiences. Are people more present in physical reality while on a rollercoaster versus the ground? In this case, mismatch between vestibular and visual senses may have meant that the high presence condition had the opposite to intended effect. Presence can also be demonstrated by looking for behaviours that spontaneously occur in the physical world, in virtual reality (Sanchez-Vives & Slater, 2005). For example, emotional reactions (Diemer et al., 2015), stress (Shiban et al., 2016; Zimmer et al., 2019), postural sway (Brookes et al., 2019; Freeman et al., 2000), the rubber hand illusion (Yuan & Steed, 2010), and fear of heights (Freeman et al., 2018; Martens et al., 2019). These demonstrations are convincing manipulation checks for the use of virtual reality to produce ecologically valid responses. Better measures of presence can have direct impact on the improvement of virtual reality as well as better research practices. For example, understanding communication between virtual reality users and people in the physical world (George et al., 2019), and the effect of interruptions on presence and data collection (Oh et al., 2019).

As a final point in this section, I want to address the question of whether people behave differently in virtual versus physical reality. This is a question I have encountered regularly in discussions with colleagues and also in the scientific literature (Biocca, 1997; Feldstein, 2019; Nilsson & Kinateder, 2015; Siegel & Kelly, 2017). For example, someone might ask whether people are less fearful in virtual reality because they know they will not really be hurt. The question is in part related to mixed interpretations of the term virtual reality. Following Steuer's (1992) definition, any differences between behaviour in physical and virtual reality are explained by limitations of vividness and interactivity. If it was clear to a user that they could really be hurt in virtual reality, as in the Ultimate Display (Sutherland, 1965), then their behaviour would adjust accordingly. However, if framed another way, the question can be used to challenge this definition of virtual reality. There are many experiences in virtual reality that are not possible in the real world that are used in the entertainment industry, as well as research (e.g. simulating being Superman; Rosenberg et al., 2013). Being able to do the impossible is one of the great benefits of virtual reality in research (de la Rosa & Breidt, 2018). One way to view this use of virtual reality is as representative of how people would behave in the physical world if they could do these impossible things, but it would be difficult to support that claim. Another feature of virtual reality is the entry into it and re-entry back into the physical world. Whether this feature is critical to the definition of virtual reality is debatable, but it may mean that people will always behave differently between the virtual and physical realities. These questions are discussed in detail in a recent review of the ethics of realism in virtual reality (Slater et al., 2020).

2.5.2. Virtual reality in research

In addition to the psychological phenomena associated with virtual reality, it has a lot to contribute to research as a tool for presenting stimuli. The technology is not yet sufficient to create the Ultimate Display (Sutherland, 1965), where every aspect of a stimulus can be carefully manipulated. However, current technology can present naturalistic stimuli, where approximation of the real world, rather than modification of its properties, is desirable. The current state of virtual reality in research is that there are many studies (20,000+ as of 2016) thinly spread across many fields (Cipresso et al., 2018). These include clinical studies, such as virtual reality exposure therapy (Carl et al., 2019) and neurorehabilitation (Schiza et al., 2019), and broad areas of basic research, such as autobiographical memory (Schöne et al., 2019), language (Tromp et al., 2018), vision and navigation (A. R. Wade et al., 2018). These closely intersect with a research from the engineering community (Abtahi et al., 2019; Cherni et al., 2020; Cipresso et al., 2018). As such, any complete review would be disjointed due to the variety of methods used (e.g. qualitative, clinical) and not of direct relevance to our research or methods. Therefore, in this section I will present several studies that show the promise of virtual reality as a research method. These studies use the unique properties of virtual reality to create manipulations and measure behaviour in new ways, independent of the specific research area.

As discussed, the definition of virtual reality puts it along a spectrum, making the judgement of whether to include a study in a review subjective. In my opinion, recent reviews of the use of virtual reality in psychology and neuroscientific research have been generous with their definition of virtual reality and the number of studies included as such (Cipresso et al., 2018; de Gelder et al., 2018; Lanier et al., 2019; Pan & Hamilton, 2018). Note, this is not the case in clinical research, where the definition is stricter. For example, reviews and meta-analyses of virtual reality exposure therapy (e.g. Carl et al., 2019; Valmaggia et al., 2016) and other clinical applications of virtual reality (Schiza et al., 2019) have been more explicit, perhaps due to being more established fields. However, the timing of these reviews does coincide with the recent step-change in technologies for virtual reality and reflect a resurgence of interest in using them for research. Until circa 2017 when the first generation of modern head-mounted displays were released, virtual reality was only accessible in specialist facilities. While those facilities could produce high quality virtual reality experiences, reduced costs and ease of operation now mean that many more researchers can feasibly use it in their research (Slater, 2018). Likely due to a combination of development time and publication cycle, there is currently a surge of new research being published that uses head-mounted displays. Below I will review research relevant to the wider topics presented in this thesis, as well as that which has contributed to the development of neuro-VR. In doing so I have set the threshold for virtual reality as involving some interaction and being separate from physical reality. By separate, I mean that senses presented to the user in virtual reality should be occluded in the physical world.

By occluding one sense and replacing it with another, researchers can study how senses interact. One classic example of this is the Rubber Hand Illusion (RHI; Botvinick & Cohen, 1998), where synchronous touching of a rubber hand (visible) and the participant's hand (occluded) results in embodiment of the rubber hand. The effect can be induced with a small amount of equipment but is difficult to extend to other senses beyond touch. As mentioned, the RHI has been replicated in virtual reality as the standard task (Yuan & Steed, 2010) and also a sixth finger variant (Hoyet et al., 2016), but it has also been developed beyond the original task to use other modalities. For example, motion capture and virtual reality can allow for movements to be transferred from a participant's body to a virtual body, resulting in Full Body Illusion (FBI; Peck et al., 2013). This research and others that have replicated the effect using camera footage, rather than virtual humans (Carey et al., 2019), have demonstrated the importance of first-person perspective in embodiment. The strength of the FBI and the relative ease of setting it up in virtual reality has allowed researchers to address more specific questions about embodiment. For example, the time course of the effect (medium-high effect after just five seconds and stable over time; Keenaghan et al., 2020), the effect of racial bias (Farmer et al., 2012; Peck et al., 2013) and congruency between pain and visual presentation of a limb (Matamala-Gomez et al., 2020). Virtual reality has also been used to demonstrate a cardio-visual Full Body Illusion (Heydrich et al., 2018). In this study, participants viewed a virtual avatar from behind which would pulse with light either in synchrony or conflict with their own heartbeat, measured and analysed using electrocardiogram in real-time. They compared the somatosensory evoked potential (SEP) from median nerve stimulation and self-reported self-identification with the virtual avatar between these conditions. Replicating their previous work (Aspell et al., 2013), they found increased reports of self-identification with the virtual avatar in the synchronous condition. They also found increased amplitude of the P45 SEP components in the synchronous condition, localised to the somatosensory cortex. This difference suggests that interoceptive signals (heartbeat) modulate the exteroceptive signals (touch) that were traditionally associated with the SEP by earlier research using the FBI and RHI. Together, these studies exemplify the use of virtual reality in research by first showing that well established effects from traditional experiments can be replicated in virtual reality and then developing our understanding of those effects.

Virtual reality methods also allow researchers to address new questions by removing problems that physical lab-based research cannot overcome. Research related to social interactions has always had difficulties controlling and recording the social interactions presented to participants (Babiloni & Astolfi, 2014; Hari et al., 2015; Pan & Hamilton, 2018; Sonkusare et al., 2019). Because of this, social psychologists have been forced to either simplify their stimuli (e.g. emotional faces; Ekman & Friesen, 1976) and maintain control or manually label events of more naturalistic stimuli (e.g. Bandura et al., 1963). However, unlike the behaviour and neural responses measured in fields like visual and auditory research, and which can be effectively manipulated in the lab, some social behaviour may only exist under natural conditions. As discussed, virtual reality can allow researchers to present naturalistic stimuli without sacrificing as much experimental control (Loomis et al., 1999; Parsons, 2015). This benefit has been demonstrated in two studies on language switching (Peeters, 2020; Tromp et al., 2018). They argue that language comprehension and related switching are dependent on social context. In the first study, they presented stimuli within a virtual restaurant. Within this context, one of 80 food types were presented on a plate in front of a virtual human who would then make a congruent or incongruent statement about the food, or an unrelated comment. Participants' EEG was analysed around these events and compared between statement types. They found higher amplitude N400 after incongruent statements, consistent with earlier findings relating N400 to semantic integration (Hagoort, 2003), thereby validating their neuro-VR approach and supporting the ecological validity of earlier findings. In the second study, a similar task was used, but to study language switching. Their aim was to investigate language switching using naturalistic cues for the switch. Traditionally, cues such as a tone or change in colour have been used (de Bruin et al., 2018; Kleinman & Gollan, 2016; Peeters & Dijkstra, 2018). In the experiment participants switched between naming objects (e.g. turtle or axe) for one of four virtual humans and their EEG and response times were recorded. Two of the virtual humans were introduced as monolingual Dutch and the other as monolingual English speakers, so participants had to switch languages depending on who they were speaking to. To enable this, all participants were bilingual Dutch and English speakers (native Dutch). Reactions times were greater when both language and listener switched, versus just listener, supporting evidence for a cost of switching language. They found no difference in ERPs between these conditions, but both showed expected N2 and late positive components (Liu et al., 2016; Peeters & Dijkstra, 2018). Arguably, this study could have been conducted without the use of virtual reality as participants' interaction was limited (they were sitting). In this instance it is unclear whether the use of virtual reality, as opposed to a standard display, allowed for more natural behaviour. A comparison of the two methods would be necessary to make this claim.

Nonetheless, these studies demonstrate how virtual reality can be used to present social interactions while maintaining control.

Like social psychology, the study of autobiographical memory is hindered by traditional research methods. Many studies of autobiographical memory have participants form memories based on stimuli presented to them in the lab (Rugg & Henson, 2005). However, differences in brain activity measured during retrieval of lab-based versus real life memories suggests that this method may be ecologically invalid (Cabeza et al., 2004). This has led to many studies of autobiographical memory relying on participant recall of events from their own lives (Cabeza & St Jacques, 2007). The problem with this method is that measures such as neuroimaging during memory formation, accuracy of recall, and age of memories cannot be made (e.g. Maguire et al., 2001). It is also more challenging to investigate forgetting (Elliott et al., 2014). For these reasons, it is essential that methods be developed to present stimuli more like the real-life experiences people form autobiographical memories in response to. This can be done by taking the lab to real life. For example, 'lifelogging' studies which monitor cameras attached to participants as they go about their daily life (Silva et al., 2018), and other similar approaches (Henkel, 2014). However, these studies are unable to control all stimuli presented to participants, especially between-subject. Further, until mobile neuroimaging techniques develop further than they already have, these quasi-experimental methods are unavailable to neuroscience. Here, virtual reality may be a more suitable alternative. Schöne et al. (2019) presented motorcycle ride experiences either with a head-mounted display or with a large monitor and tested participants' memory of scenes from the ride. They found that the head-mounted display condition caused retrieval performance to double, versus the monitor condition. Work from the same group has shown that experiences presented in virtual reality are more vividly retrieved (Kisker et al., 2019). These studies show how research into fundamentally naturalistic phenomena, like autobiographical memory, can be investigated under controlled conditions.

The experiments described so far focus on what is displayed to the user, but researchers can also take advantage of some of the technologies inherent in virtual reality that are not directly related to presentation. Many of the specialist technologies used for virtual reality measure the user's behaviour (e.g. movement, such as walking or pressing buttons) and translate it into the virtual reality (e.g. moving through the environment, picking up objects). Normally, this is done with great care to map the actions one-for-one to improve telepresence (Slater et al., 1995). In most research, better mapping is desirable and for some, such as the FBI work presented earlier, it is essential. Note, for experiments based in the physical world, perfect mapping is the only option, with the exception of illusions caused by careful manipulation (e.g. with mirrors; Tajima et al., 2015). However, in virtual reality, it is more feasible to manipulate mapping, so long as it was accurate before the manipulation. These manipulations can be total, affecting all senses, or selective, affecting one or part of one sense with input in virtual reality. For example, one study took advantage of room-scale motion tracking combined with a head-mounted display to manipulate how users' walking movement was translated into the virtual world (Abtahi et al., 2019). While their investigation was focused on improving user experience, these kinds of manipulations could feasibly be used to investigate visual flow and vestibular integration. Traditionally these studies have had to be conducted using treadmills with large screens or walkways with stimuli projected along them (Ludwig et al., 2018; Prokop et al., 1997). Using virtual reality methods may improve on these methods by removing confounds such as static surroundings.

2.6. Guide to remaining chapters

In this literature review I have presented and discussed research related to the study of police decision making that guided our own research. In Chapters 3-5, I will add to this review with research related to the implementation of methods and analysis. I will also present research that specifically informed the hypotheses of each experiment in more detail. Later, in Chapter 6, I will interpret the findings from our experiments within the context of all the above.

Chapter 3: Developing Methods for Neuro-VR

3.1 Introduction

The main purpose of this chapter is to present the decisions we made about how to use virtual reality as a research method in neuroscience, and the reasoning behind them. Using these methods is time consuming and requires substantial resources, so we wanted to use the same methods in all experiments. Therefore, it was essential that choices made at this early stage about virtual reality technology and how to implement it would allow us to meet the widest possible requirements. This meant developing general pipelines for creating resources for experiments, including animations, virtual humans and environments. It also meant creating plugins for software to add essential functionality for running experiments: precise timings, and communication with external hardware and other software. With these tools we could design our experiments while contributing to the development of neuro-VR as a research method.

Throughout this work, the impact of virtual reality on the quality of EEG and behavioural data was considered. The benefits to ecological validity and experimental control afforded by neuro-VR (Sonkusare et al., 2019) would mean very little without accurate, reliable data to back it up. In addition, because we intended on using the same methods in all experiments, it was essential that the choices made at this stage were suitable for all our planned research. Fortunately, there are many sources of guidance for the optimal recording and pre-processing of EEG for research (Hari & Puce, 2017). In addition, we were able to follow more specific advice about the use of EEG in more naturalistic settings (Gramann et al., 2014). As discussed in Chapter 2, neuro-VR is fast becoming a popular research tool due to increased accessibility of the technology. However, little information is available about the optimal integration of virtual reality and EEG. Providing this information for other researchers who are currently using, or would like to use, neuro-VR methods in their work is the additional objective of this chapter.

In addition to considerations of data quality, neuro-VR presents new challenges that researchers with traditional neuroscience training may need to consider. In particular, researchers need to develop new skills and/or work in a more interdisciplinary way with computer scientists and game developers

(A. R. Wade et al., 2018). Good understanding of the current limitations of virtual reality technology is essential for setting expectations of what can be done. There should be a distinction between research that aims to advance virtual reality and research that intends to use it. Knowledge of the most difficult challenges of virtual reality is essential in the latter case, so that they can be avoided, and research can progress without having to first tackle them. Time and resources available for the development of neuro-VR experiments can then be better allocated. With this in mind, this chapter will also go into some detail about the technology we used to present virtual reality and create simulations for use in all experiments. I have included only details that are pertinent to understanding the effects of using the technology on experimental design and analysis. Terms introduced here will be referenced throughout the remainder of this thesis, but with less detail, so this chapter will serve as a guide for the others.

3.2 Virtual reality options

As discussed in Chapter 2, virtual reality is not defined by the technology used, but rather the effect of using it (Steuer, 1992). However, there are advantages and disadvantages of each technology that are pertinent to research. Because of this, one of the first decisions we had to make for our project was which virtual reality technology to use. Many different technologies have been used for neuro-VR research, and there is no clear standard approach. This meant we were not limited by the compatibility of a pre-made toolbox for creating experiments in virtual reality and so we could choose our solution based on merit for our specific project. Therefore, we evaluated techniques based on how well they would integrate with EEG (and later, OP-MEG) and our experimental protocols.

One option was to use a cave automated virtual environment (CAVE: Cruz-Neira et al., 1992) system. A CAVE surrounds the user, partially or completely, with a display of a virtual environment. There are many ways to do this, but most commonly CAVEs are formed of stereoscopic, back-projected display walls and a floor and/or ceiling display (Muhanna, 2015). Directional audio may also be used. Usually the image and audio can be updated not just based on the content of the virtual reality, but by tracking the user’s head position and orientation as they move through it. Due to difficulties in updating images and audio to be viewed/heard from two different perspectives, most CAVEs are confined to a single user, although multi-user setups are possible (Kuchera-Morin et al., 2014). In an ideal CAVE setup, from the perspective of the user the display walls are not easily visible and appear as a unified three-dimensional view of a virtual environment. The walls are nonetheless present and limit physical movement, so a warning system based on proximity to the walls is advisable. To interact with the environment, a 'wand' controller can be used (Muhanna, 2015). The wand motion is tracked by the same motion capture system as the user's head, so its position relative to the user and object in the virtual environment is known. This means that the user can point and grab objects, navigate a graphical user interface and record their responses.

We strongly considered using a CAVE because of its compatibility with EEG. Physical integration of CAVE hardware with an EEG cap would be simple, as most CAVE hardware is not attached to the user. Typically, only a pair of glasses (to allow for stereoscopic viewing and motion capture) are required, and these can easily be worn over an EEG cap. Most systems use 'active' glasses, which contain electronics which shutter or block each eye interchangeably, in sync with the display (DeFanti et al., 2011). While it is possible that these electronics could cause noise in EEG recordings, this would only be in higher frequencies of around 60Hz or higher, as this is the standard rate the shutters operate at (Török et al., 2014). However, the glasses can also be entirely passive by using infra-red reflective markers for motion capture and polarised lenses for stereoscopic vision (DeFanti et al., 2011). The effect of this is that electronic interference from a CAVE to EEG is minimal and outside frequencies of interest (Török et al., 2014). A CAVE is also unique in that the user can see their own body within the virtual world, just as they do in the real one (Cruz-Neira et al., 1992). This should contribute to a feeling of presence and so is beneficial to research, unless an experiment requires manipulation of the user's body in virtual reality (for example, for body size illusions; Weber et al., 2019). Despite these benefits, we decided not to use a CAVE for the presentation of stimuli in our experiments. Due to the number of displays and their high resolution, the graphical processing power required to present a scene on them is large, relative to what is required to render a scene to a standard, single monitor display (DeFanti et al., 2011). This means that even using modern, high-end hardware, we would have had to use simple environments, or create them in an extremely efficient way. This was in opposition to the requirements of our project, as we needed to present realistic scenes and virtual humans, and we lacked the resources to commit to ensuring efficiency of the simulations. Further, CAVEs are not common and so the software used to create applications for them requires specialist knowledge. This means that there is less support for developing applications, or, in our case, experiments. Logistically, a CAVE was also impractical as it is a fixed piece of equipment, meaning data collection could only have been done at sites with a CAVE. Limited access to CAVEs also means that studies are more difficult to replicate for other researchers. For these reasons, we sought alternative virtual reality technologies.

As mentioned in the introduction section of this thesis, modern virtual reality head-mounted displays were becoming readily available around the time we were planning our experiments. We evaluated these against a CAVE and decided to use them in all experiments. Just as for CAVE systems, there are many different types of head-mounted display, with different features we had to evaluate for use in our experiments. Generally, head-mounted displays are aptly named: they are a display which covers the user's eyes like goggles, blocking out the real world and presenting a virtual one. Just as for CAVE systems, the display technology can vary, but typically they contain a lens and display for each eye. The lenses allow the wearer to focus on the screens comfortably, despite their proximity, and make them appear larger to provide a greater field of view. Like the CAVE system, the image on the display is updated based on the user's movement. The effect of this is that the user can view the virtual environment in a similar way to how they view the real world. However, less of the virtual environment needs to be rendered when using a head-mounted display than a CAVE, because anything the user is not looking at is off-screen. This means that the virtual environments displayed using a head-mounted display can typically be made more complicated, as the resolution is more comparable to that of an ordinary monitor. Another advantage of this is that head-mounted displays require very little specialist software, as they are similar enough to the hardware that developers have been working with for decades. Further, plugins to enable functionality in commonly used development software exist for most head-mounted displays as they are consumer products. In our case, and likely for many other researchers, game development was not our main skillset and so being able to rely on the support of larger communities was important.

Despite recent advances, head-mounted displays still have some major limitations. This is in part due to the requirement to minimise motion-to-photon latency (MTPL) –a measure of the time between a user's movement and the display updating according to this movement. One often cited figure for this maximum MTPL comes from mixed industry sources and is usually 20-60ms (Abrash, 2012; Lawlor, 2016), but this data is not published. Nonetheless, manufacturers of head-mounted displays have focused their product development on reducing MTPL. The displays must operate consistently at high refresh rates too because this determines the minimum motion-to-photon latency. For example, at 90Hz (used by the Oculus Rift CV1 [Facebook, Inc., Menlo Park, CA, USA]), even with no latency introduced by other processes, the minimum MTPL would be 11.1ms. To ensure that this can be achieved by even high-end hardware, the resolution of the displays must be kept relatively low, at around the same resolution as computer monitors. However, the display in a head-mounted display must provide a large field of view and therefore looks more pixelated than a monitor. This pixelated effect is enhanced by magnification of the gaps between pixels, which can create a mesh-like appearance and result in the 'screen door effect'. Because of these limitations, any application that requires observation of small features, such as faces or text, is not suitable with current generation head-mounted displays. Until head-mounted displays can use higher resolutions, experiments presented using them must consider their impact on visual acuity distance of stimuli presented to participants. Over the course of this project (2017-2020) head-mounted display technology has improved, but no step change has been had in graphical processing power.

For applications that can accept these caveats to performance, there are some considerations that are more specific to research. To assess these, it is important to understand more about the technology used in head-mounted displays. One consideration is how the head-mounted display receives its signal. With a few exceptions, high-performance head-mounted displays require a physical link between a computer and the display to send and receive display information and motion capture data, respectively, as well as provide power. For wireless systems, a battery and power supply must be attached to the device and a wireless transmitter used to communicate with the computer. These physical setups both introduce problems. While cables may be several metres long, even with careful cable management, user movement (and particularly rotation) is still limited. Further, having a cable attached to the user's head is likely to introduce a large electrical line noise artefact into the EEG recording, unless optical isolation is used. The same is true of wireless systems, as there is still a connection between the battery and the user. Wireless transmitters operate at very high frequencies and low amplitudes, so do not interfere with EEG, but their power supply would likely introduce artefacts.

Another important consideration is the motion capture system used by the head-mounted display. To avoid motion sickness, it is important that head-mounted displays can track translation and rotation, known as six degrees of freedom (6DOF) motion capture (Gourlay & Held, 2017). This needs to be achieved with low latency, because it is the first stage in reducing MTPL. To do this, most displays rely on a fusion of internal sensors, including gyroscopes, magnetometers and accelerometers. In the Oculus Rift CV1 these sensors sample at 1000Hz (Lavalle et al., 2014), meaning they contribute just 1ms to total MTPL. From a starting reference point this fusion of sensors can provide precise tracking for short periods of time, after which they begin to drift without external reference (Kok et al., 2017). There are several methods for solving this problem, and this is one of the ways in which different head-mounted displays vary the most. The main distinction is between outside-in and inside-out tracking. Outside-in tracking uses external sensors to look inside the tracking space (Gourlay & Held, 2017). For example, the Oculus Rift CV1 head-mounted display has infra-red light-emitting diodes (IRLED) in fixed positions across its surface which can be detected by external sensors in a system known as the Constellation (Lavalle et al., 2014). Inside-out tracking relies on sensors placed on the head-mounted display to look out into the tracking space and understand its position relative to points within it. The two common methods of inside-out tracking are the Lighthouse system (Valve Corporation, Bellevue, WA, USA) and the use of structured light sensors that identify the positions of fixed objects in the environment and use them as references for the internal sensors. The difference in performance between these methods is minimal and so does not directly affect experimental protocols. However, they likely affect EEG recordings by introducing complex electrical artefacts (G. Roberts et al., 2019). For example, the IRLEDs used in the Oculus CV1 are powered by three controllers drawing a relatively large amount of power (iFIXIT, 2016). Generally, it can be assumed that the more electrical equipment attached to the head-mounted display, the greater the power and complexity of noise introduced to the EEG recordings.

Comfort was another factor that we needed to consider when creating protocols for our virtual reality experiments. It is important not just because participants' wellbeing is important, but because discomfort may remind them that they are not really in the virtual environment and this may detract from their feeling of presence within it. Relative to a CAVE, a head-mounted display involves the user wearing a lot more equipment and this can be uncomfortable. The display itself can also cause some discomfort. As mentioned earlier, the displays used are made to appear very large, to allow a wide field of view. This is done by having the wearer view them through lenses. The standard for current generation head-mounted displays is to use a lens with a fixed focal length, typically around two metres. This means that when viewing objects in the virtual environment that are not two metres away, the wearer's eyes are focused at the wrong distance. The 'zone of comfort' for viewing with a fixed focal length of two metres is between one and five metres (Shibata et al., 2011). Presentation within this range can still cause discomfort over time. Therefore, research using head-mounted displays to display stimuli should try to match its distance to the focal length of the lens. The lenses can cause other issues as well, as most manufacturers take a 'one size fits all' approach and have the lenses fixed to the display. This means that people who require glasses to correct their vision must wear them as well as the head-mounted display. This can be uncomfortable, or even impossible, if the glasses are large.

Having considered these factors, we decided to use an Oculus CV1 head-mounted display. The only other option from the first generation of consumer head-mounted displays was the HTC Vive. There is very little difference in performance between the two devices, but we determined the physical setup of the Oculus Rift CV1 to be easier to integrate with EEG. The straps on the Oculus Rift CV1 were easily adjustable to go over an EEG cap, as well as washable. In comparison, the HTC Vive had a rigid headband with a hard to clean sponge cover. The HTC Vive is also known to have issues with water damage, making cleaning riskier. Additionally, the Oculus Rift CV1 allows for adjustment of the space between the lens for each eye to match the wearer's inter-pupillary distance, improving comfort and reducing the likelihood of having to reject a potential participant in our study due to incompatibility of the display.

3.3 Using a game engine for neuroscience

Software is equally as important as hardware to the integration of virtual reality with neuroscience. Bespoke programs and traditional psychology software, such as The Psychophysics Toolbox (Brainard, 1997), can be used to present stimuli in virtual reality (e.g. Fulvio & Rokers, 2017), but there are more powerful tools available outside of research, in the entertainment industry. As discussed in the introduction to this thesis, we believe that these tools for creating high-fidelity simulations can be used to great effect in cognitive neuroscience. We used software called Unreal Engine 4 (UE4; Epic Games

Inc., Cary, NC, USA) to program and present the simulations used in all experiments. UE4 is a free (with licensing terms) and highly developed game engine that provides a development environment of integrated packages normally used for making games for consoles, computers, mobiles, and virtual reality. These games are often made by groups of people who specialise in different areas of game design, such as animation or lighting. Fortunately, this is not a requirement for all projects. The game engine software is designed so that small teams and independent developers can use it with limited specialist knowledge, as is likely the case for most researchers.

A great deal of support is available for new users of UE4, including built-in tutorials, video tutorials and active forums. Additionally, there is a virtual marketplace with useful assets for creating games (and experiments). These include components like buildings and materials to create environments, animations, and sounds. This further helps independent developers who do not have the resources to make bespoke assets. The main alternative to UE4 is Unity (Unity Technologies, San Francisco, CA, USA), which provides similar functionality. Unity is arguably now the more popular game engine for neuro-VR, and some groups are working on toolboxes to help researchers create experiments with it more easily (Bebko & Troje, 2020; Brookes et al., 2019). The defining difference between the two toolboxes is that UE4 is based in C++ and has built-in visual scripting, called Blueprints, and Unity is based in C# and uses traditional coding for programming. We decided to use UE4 because it has a larger marketplace for assets, and we felt it easier to approach due to the option of using Blueprints for scripting.

UE4 contains many useful built-in functions to help developers make games but lacks a lot of the standard functionality normally required to create experiments. This is understandable but highlights why there is a need for toolboxes and sharing these functions between research groups. Fortunately, as UE4 is based in C++, it is possible to create functions which can be used in UE4 to perform tasks outside of it. Highlighting the excellent community of UE4 users, we often used guides to help do this. While we created many useful functions for making and running experiments, the two most likely to be needed by researchers are the ability to read and write text files and communicate with external specialist hardware. We found that reading in text files was essential for inputting trial information that had been randomised and counterbalanced in external software into the experiment. Writing text files is likewise important for recording logs with information about what happened in trials, required for behavioural data analysis. To create this functionality, we adapted a tutorial to our needs (pally qle, 2016). In our case, the specialist hardware we needed to communicate with was the EEG amplifier, in order to send event markers to guide event related analysis. The standard method to do this is by sending signals from a parallel port card, as this introduces minimal latency and, most importantly, minimal jitter. Here, jitter refers to the reliability of the latency, which is particularly important as, unlike a known latency, it cannot be corrected. Unfortunately, parallel port cards are obsolete for most applications and so are not widely supported (certainly not in UE4). We used legacy drivers maintained by Logix4U and Phillip Gibbons by adding a function library they produced to a custom UE4 plugin. This essential step allowed us to address the parallel port and send signals to the EEG amplifier directly from Blueprints in UE4. An alternative method would have been to use an interface such as Lab Streaming Layer to communicate between UE4 and the EEG computer (Kothe et al., 2018/2020).

3.4. Development pipeline for scenarios

Much of the work to create experiments using a game engine was closer to game design than neuroscience. To understand the effect of this, and how it differs from other methods, it is important to know more about the development pipelines involved (for an overview, see Figure 5). This is especially true for anyone working on a neuro-VR project, whether they are working on development directly, or in collaboration with others. Knowing the strength and limitations of development tools can and should feed into experimental design from an early stage. The main pipelines were animation and environment creation, which were combined in the game engine with added functionality to form complete scenarios. It is important to note that development of these pipelines was continuous and in parallel to other aspects of the research project, so improvements were continuously incorporated.

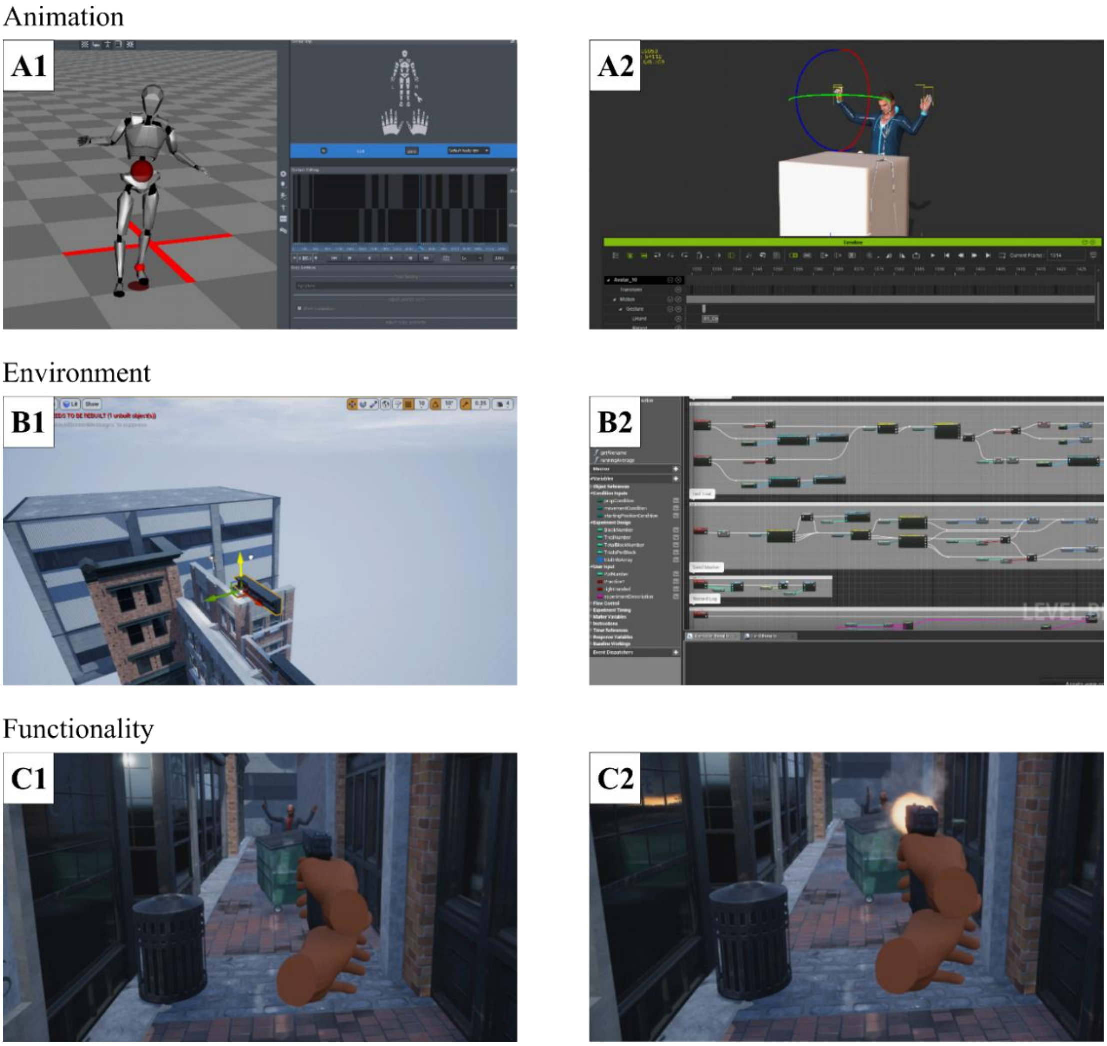


Figure 5. Overview of development pipelines used to create our virtual reality experiments. The top panels show our animation pipeline: A1) We used our motion capture facilities to record full body movements of an actor. This allowed us to create highly realistic animations. A2) We then used animation software to remove any errors in recording and then added the animation to a virtual human, ready for import into the game engine. The middle panels show how environments were made for scenarios: B1) First, we created a ‘level’ for the scenario. To do this, we used purchased 3D assets, such as modular buildings. We then added the virtual human and starting position for the participant. B2) Next, we used Blueprints to allow for the use of props and lighting and to give the experiment control over animation timing. The bottom panels show how animation and environment were combined into a functional scenario: C1) The animation was set up to play out in front of the participant. In this case, the virtual human has picked up a pistol and is complying. C2) The virtual human then attempts to fire at the participant, who fires first. The animation then blends into a ‘shot’ animation in which the virtual human falls down.

Within UE4, animations involve, at a minimum, three classes of object. These are the skeleton, skeletal mesh and animation (see Figure 6). Skeleton objects describe the layout of the joints and how the bones between them rotate to cause constrained movement (Naour et al., 2019). For the animations used in this project, the skeleton resembled a simplified humanoid skeleton, matching a standard within UE4. The hierarchy within the skeleton is non-interchangeable as, using a technique called forward kinematics, all motion from the joint or joints above the current position in the hierarchy is additive (Boulic et al., 1995). For example, if the right shoulder rotates but the elbow does not, then no position data is updated for the elbow–it just moves with the shoulder. Of importance is a bone called the root bone; this root sits at the top of the skeleton hierarchy and dictates the position, within space, of the whole skeleton. If the root bone is animated, then the character may move through space accordingly, in what is called root motion. The skeletal mesh is bound to the skeleton to give it shape. This mesh is comprised of many polygons which stretch according to their bindings with the inherited skeleton (Naour et al., 2019). A texture can be wrapped around the mesh to give it a realistic appearance. For instance, the appearance of a clothed, adult human male. Usefully, the skeletal mesh for a given skeleton can be varied without affecting anything else which is what enables the use of a different character for each scenario. The final object class drives the movement of the skeleton, the animation sequence. This acts on the skeleton within the forward kinematic framework previously described. Animations take the form of keyframes played at a constant rate. For each keyframe, values are associated with joints describing their current position in relation to the joint above. UE4 can blend between two or more animation sequences to create a smooth transition between them. These can be created on-line to great effect. For example, at any point during an animation, if the virtual human is shot then their animation can blend into a more suitable one, such as the virtual human falling to the ground. Separate animations may also be applied to different sections of the skeleton object, such as a walking animation used to control the legs of a virtual human, but a surrendering animation for the arms.

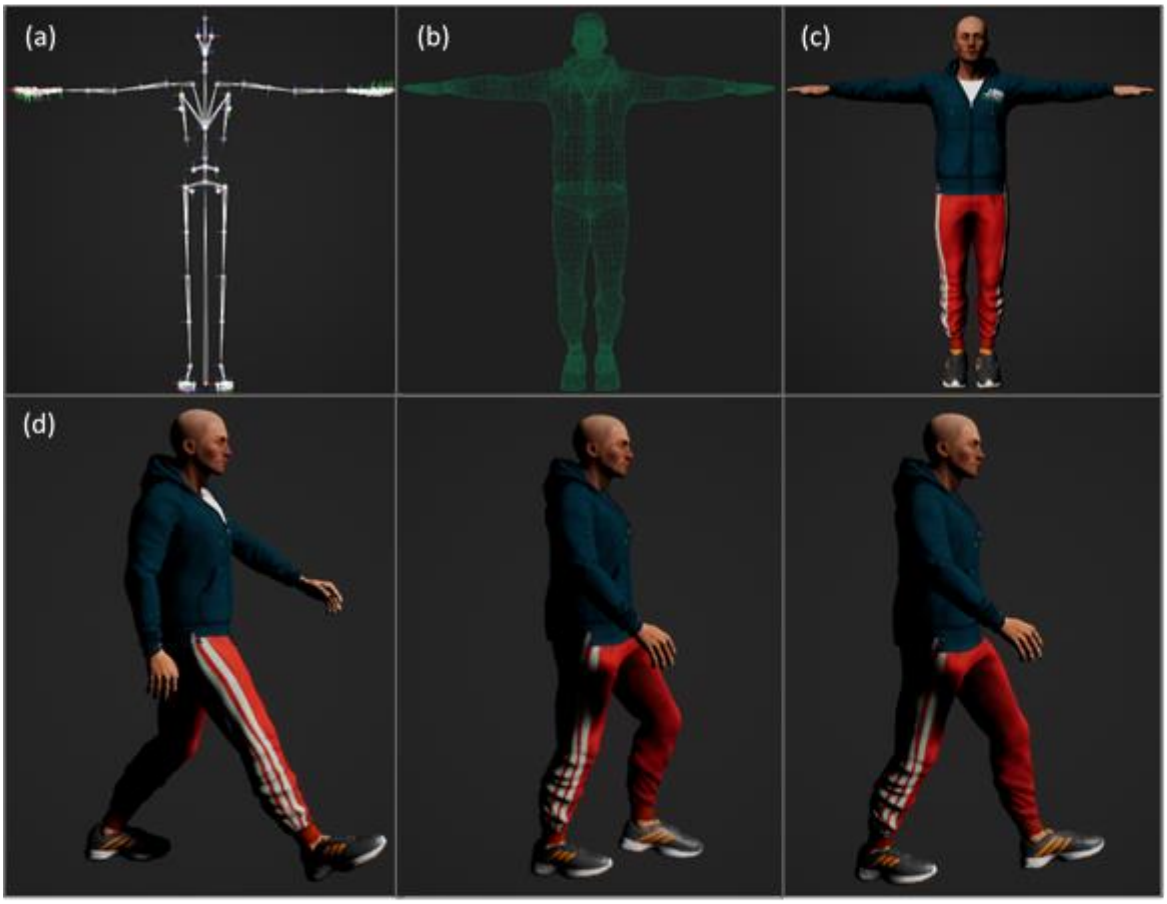


Figure 6. The three classes of animation object within Unreal Engine 4: (a) A humanoid skeleton made up of joints and bones. The root bone can be seen at the base of the skeleton, between the two feet. (b) An example of a skeletal mesh made of polygons. (c) A texture wrapped skeletal mesh. (d) Three parts of a walking animation sequence.

While UE4 has some animation functionality, for most applications external software and hardware is required. To create the animations used in all experiments presented in this thesis, the following development pipeline was used. Virtual humans were created using a software called Character Creator 2 (Reallusion Inc., Taipei, Taiwan). Within Character Creator 2, we created realistic virtual humans using resources within the software, such as clothing, body shapes and materials. These were then exported directly into UE4, where only minor adjustments were required to make them compatible. Animations for these virtual humans were first recorded using a Perception Neuron 32 Alum motion capture suit and software called Axis Neuron Pro (Noitom Ltd., Beijing, China). The body suit streamed the position of 32 points on an actor who moved in the way intended for the animation. The sensors that record the positions work in the same way as most head-mounted display motion capture, by using sensor fusion. However, they lack external reference, so only short animations could be recorded before drift occurred. For this reason, animations required some ‘tidying up’ to make the motions look natural. To do this, we transferred the animations to iClone 7 (Reallusion Inc., Taipei, Taiwan) using a plugin to map them directly onto the same skeleton used for the virtual humans. This is referred to as retargeting, which, without an automatic process, can be painstaking. The animations could then be edited using a combination of techniques, including inverse kinematics (Wilhelms, 1987), animation blending and keyframe-by-keyframe animation. A final step was to export the animation to the 3D creation software, Blender (Blender Foundation, Amsterdam, Netherlands), to make the animation compatible with UE4. This required adding root bone animation by extracting changes in position of the centre of the animation. From there, the animation sequence could be exported from Blender and imported into UE4, where it could be used to drive the animation of virtual humans within scenarios.

One of the main purposes of a game engine is level design, so most of the creation of 3D environments was done in UE4. As mentioned earlier, there are many pre-made assets available for free or at relatively low cost. Normally, some time would be spent by developers on creating the 3D assets to populate an environment, because this allows for consistent artistic style. For industries such as game design this is essential, but, for researchers, time may be more valuable than aesthetic. We purchased assets including buildings, ground materials, and weapon models that met our requirements. Our only bespoke 3D artwork was to take apart some of the weapon models to allow for independent animation of some components (for example, the trigger). This was done using Blender software (Blender Foundation, Amsterdam, Netherlands). With these assets, we constructed the virtual environments. To do this, 3D objects were first arranged in realistic positions and with suitable scale. Properties were then assigned to these objects which described the way they could interact with the world. For example, how lighting affected them (e.g. casting shadows) or whether they could be walked on or moved. We found that a few important considerations needed to be made for creating environments for virtual reality. As mentioned, a requirement of high-quality virtual reality is low-latency and high-framerates. To achieve this performance, the resources required to render the environment must be assigned with care. For example, dynamic lighting, such as moving shadows, is very costly and should only be used for important features of the environment, if at all. Resources can be saved by using lower quality textures, because head-mounted displays do not have the resolution required to visualise high quality ones.

3.5. User interface

Using a head-mounted display, participants can interact with the environment in two ways. First, they can move through the environment which affects what is presented to the display. Primarily this is achieved by 6DOF tracking of head position, but also of hand controllers which can be viewed as 3D models of hands, or an object held by the user. In the case of the Oculus Rift CV1, hand position is tracked using an Oculus Touch controller held in each of the user's hands. These controllers benefit from a few touch sensitive surfaces which can be used to interpret the way the controller is being held. We used this functionality to drive animations of realistic human hands animated in Blender (Blender Foundation, Amsterdam, Netherlands). This was achieved by interpreting hand position from the input combination of interfaces on the controller and blending between them as they change (Figure 7), with the assumption that the participant held the controller correctly. While not essential for our experiments, these extra animations improved the mapping and range of actions available to participants (Slater, 2009; Steuer, 1992), possibly increasing their feeling of presence in the virtual environment.

The second method of interaction is more obvious and is the use of the physical buttons and triggers on the controllers. For our experiments, and likely many others, these are the primary way of recording participants' responses and allowing them to affect the environment. When discussing their use in experiments, we had some concern about whether individual differences between participants would affect performance when using these controllers. Certainly, it is true that familiarity with some entertainment platforms, particularly gaming consoles, is essential for using joysticks to navigate 3D environments and unrestricted use of controllers with multiple buttons. However, this stems from poor action mapping–if the use of these controllers were more similar to real world use then first-time users would already be experts. That is not to say that the Oculus Touch controllers have perfect mapping for all their functions, but for actions such as grabbing (middle finger trigger), firing a weapon (index finger trigger) and general movement of the hand, the actions are quite natural. The joysticks and multiple buttons on the top of the controllers are less well mapped. In our experiments we decided to disable the joysticks and map the remaining buttons ('A' and 'B', 'X' and 'Y') to the same action per hand.

Anecdotally, of the 100+ participants tested in our experiments, all rapidly mastered the use of the controllers.

Figure 7. Screen captures of hand renderings within the game engine. The top panel shows the hand positions presented within the game engine and the bottom panel shows how the controller was held to produce them. This is a sample of three from ten possible hand positions that can be interpreted.

3.6. Recording events

Most EEG analyses benefit from precise timings for when discrete events, such as stimulus onset or response, occur in the otherwise continuous data (Puce & Hämäläinen, 2017). The degree of precision required is dependent on the type of analysis. For example, a resting state analysis with two, five-minute-long conditions, ‘eyes closed’ and ‘viewing fixation cross’, would only need to know approximately when each condition started as the beginning and end could be easily removed. However, analysis of ERPs in time-series data and changes in power in time-frequency data demand more precise timings. This is clearest for ERPs because they are interpreted in tens of milliseconds (e.g. the N170; Bentin et al., 1996). This means that if an event is recorded too early or too late (and this is not corrected for) then the interpretation may be wrong. ERP analysis is also susceptible to jitter as it relies on the ‘averaging out’ of noise from the time-series. If there is jitter then the signal, in part, is removed as noise and the resulting ERP smoothed and reduced in amplitude. The same is true for time-frequency data, but to a far lesser extent. Time-frequency data is often presented with high temporal resolution because that is useful for interpretation of the observed effect from averaging. However, we do not actually observe changes in power at high-temporal precision around individual discrete events in EEG. This is certainly true for lower frequencies. As will be described in more detail in the results section of Chapter 4, the availability of high-temporal resolution average effects comes from interpretation of the change in power within much wider time windows as representative of the centre of that time window. The resolution of that central time window is arbitrary, but values from 20 to 50ms are often used. Precise timings are nonetheless desirable for time-frequency analysis, but less relevant given the much greater reduction in temporal precision versus time-series analysis.

Experimental software packages and toolboxes consider how they can reduce latency and jitter when recording event timings and sending triggers to external devices, such as EEG amplifiers. The performance of different software has been compared in a recent review (Bridges et al., 2020), showing that most lab-based software (i.e. not web-based) perform well enough for EEG analysis. Game development software like UE4 and Unity were not included in their review as they are not traditional experimental software. However, recent interest in neuro-VR has motivated others to test them. Experiments created in UE4 and presented on an HTC Vive head-mounted display have been tested (Wiesing et al., 2019). They found that, in one experiment, the timings were reliable (sub millisecond precision), but that in another they were not (12ms precision) when using the internal timing of the game engine. This suggests that differences between experiments may affect timings. For example, loading or saving data such as log files or 3D assets can cause unexpected delays and should be avoided. When creating our experiments, I ensured that files were only read and written during intertrial periods to avoid this. Internal tests of timing precision (asking the game engine to record a timestamp every second while running different tasks) suggest that this is effective. Others have reported similar allowances using the Unity game engine (Brookes et al., 2019) and this is also common practice in the games industry. These timings are only related to what the game engine intends to present on the head-mounted display. I was not able to test the delay between a stimulus onset and its detection on a photodiode. Fortunately, the low and consistent MTPL described earlier ensures minimal delay. It is also continuously measured by the device and the log files show consistent 27ms MTPL. The stability is supported by tests of the HTC Vive Pro using both Python and Unity to present stimuli which found high precision and consistent timing offset related to MTPL (Chénéchal & Chatel-Goldman, 2018). Relative to experimental software, game engines are not as precise and so researchers who are running experiments which demand absolute precision must take extra care to measure the precision of the virtual reality setup and may need to make specific allowances. Promisingly, the Python-based experimental software, PsychoPi (Peirce et al., 2019), has recently released functionality for presenting experiments in virtual reality. This may be the solution for researchers with those priorities.

A final, and crucial, consideration of event timing in game engines is the type of stimuli. So far, the consideration has been for discrete and measurable events such as a flash of light or a response button being pressed. However, the continuous properties of naturalistic stimuli mean that timing precision is less important. Rarely in naturalistic stimuli does a stimulus onset immediately and come from nothing in the same way that a tone or flash of light does. For example, transitioning between standing and walking or speaking a word as part of a sentence. In a game engine the timing of these events is known because they are initiated by it and so a precise time can be recorded. However, this time should not be overly relied on as it does not necessarily reflect the time the transition was observable. Some stimuli timing can be recorded more precisely. For example, you can record when an object appears on screen by checking if the camera view in the game engine has line of sight to that object, but even that may be imprecise. How do you decide when an object has been presented if it appears gradually around a corner? And how does that compare to the presentation of another object? For naturalistic stimuli the only measurement that can be taken as a ground truth is the response. While important measurements, such as reaction time, are dependent on valid and precise stimulus times, EEG data is less dependent on them. For this reason, we designed our experiments for response-locked EEG analysis.

3.7. Considerations for EEG recording

Good experimental design is essential to maximise the signal-to-noise ratio of EEG data. Additional measures must be taken to ensure that EEG recordings do not contain artefacts, or that these are minimised. Here, artefacts are defined as sources of noise not naturally present in the phenomena you intend to observe (Hari & Puce, 2017; Teplan, 2002), and which have likely been introduced by the experimental protocol. As the aim of EEG is to record electrophysiological data from the brain, any other source of electrical activity recorded in EEG can be considered artefactual. These sources commonly include electrical devices, movement of sensors, and muscle stimulation (Hari & Puce, 2017). Muscles near the sensors on the scalp are particularly problematic, such as extraocular muscles which present clearly over frontal and temporal electrodes. Artefacts may also include measured brain activity unrelated to the intended experimental manipulation. For example, those arising from distractions in the task or mind wandering (Braboszcz & Delorme, 2011). Introduction of some artefacts is inevitable in a neuro-VR or naturalistic stimuli-based experiment (Klug & Gramann, 2020). This does not necessarily condemn the approach, as most EEG experiments are affected by unavoidable artefacts; it is very challenging to stop people blinking or moving their eyes and even more so to stop electrocardiographic artefacts.

Fortunately, methods for removing artefacts exist and will be described in detail in the next section, as well as in the other empirical chapters of this thesis. Briefly, these include methods for identifying unique sources in multi-channel data and statistical methods (averaging and inferential statistics). However, best practice is still to avoid introducing them in the first instance (Teplan, 2002). Artefacts can be minimised by avoiding participant/cap movement and proximity to electrical devices. Clearly, in our experiments, this was not going to be possible, but practical steps could be taken to reduce the impact of participants standing while wearing the head-mounted display. One way we did this was to design scenarios that presented all important stimuli in front of the participants and so did not require much movement or looking around, thereby avoiding motion artefacts (Tremmel et al., 2019). We were also concerned with possible movement of the sensors due to contact with the head-mounted display, which others have reported since (Nenna et al., 2020). This was largely mitigated by proper adjustment of the straps by the experimenter so that they were tight with the cap. Participants were never allowed to put the display on over the EEG cap themselves.

Other artefacts were still introduced by the head-mounted display. We ran some pilot tests to show the differences between wearing it and not to determine whether electrical sources were above frequencies of interest, as others have found for their own experimental setups (Hertweck et al., 2019; Török et al., 2014). For this thesis I have supplemented the data from those tests with larger datasets. For the ‘head-mounted display on’ dataset I have used continuous data from the experiments reported in this thesis. For the ‘head-mounted display off’ dataset, I acquired some continuous data from a colleague who collected data in the same facility and using the same EEG equipment and setup, but with a more traditional, computer-based task. The tasks varied greatly but were suitable for these non-event related comparisons. Power spectral density was calculated between zero and 250Hz from the continuous EEG for all individuals and all channels. It was then averaged across individuals, but not channels, to allow inspection of the topographies of identified peaks. Figure 8 shows the results of this analysis.

We had two planned comparisons between datasets. The first was 50Hz line noise, which was likely to be present in both datasets, as neither were recorded in a Faraday cage. The line noise in the head-mounted display recordings was found to be several orders of magnitude greater than for the computer-based recordings. The power supply for the head-mounted display was five volts direct current (DC) supplied by a pin on the USB connection to the computer used to render the virtual environment. While this would not have caused 50Hz noise in itself, the proximity of the cables to the 50Hz power supply of the computer may have. Magnetic fields generated by the power supply induce current in surrounding conductors, increasingly for proximate, highly conductive materials like copper cables. This is why there is some line noise present in the non- head-mounted display, only to a much lesser extent (Ferree et al., 2001). The second planned comparison was at 90Hz, the refresh rate of the displays inside the head-mounted display we used. As expected, we observed greater power at 90Hz on the dataset when in the head-mounted display. Because it was well above our frequencies of interest and small, relative to line noise, this was not a significant issue for data analysis (Hertweck et al., 2019).

Aside from these two well documented frequencies of 50Hz and 90Hz, we were uncertain what frequencies other components inside the head-mounted display would produce. Most notably, we observed large peaks at 52.1Hz and its harmonics. These peaks were lower in power than line noise, but still orders of magnitude greater than recordings without a head-mounted display. The topography of these peaks suggests the source may be the headband, which contained electronics for IRLEDs, which we know consume a relatively large amount of power (iFIXIT, 2016). In our work integrating OP-MEG with a similar head-mounted display, the Oculus Rift DK2, we found a similar pattern of interference which could not be explained as line noise harmonics or refresh rate (G. Roberts et al., 2019). However, without further investigation, outside the scope of this project, we can only speculate about the cause of this noise.

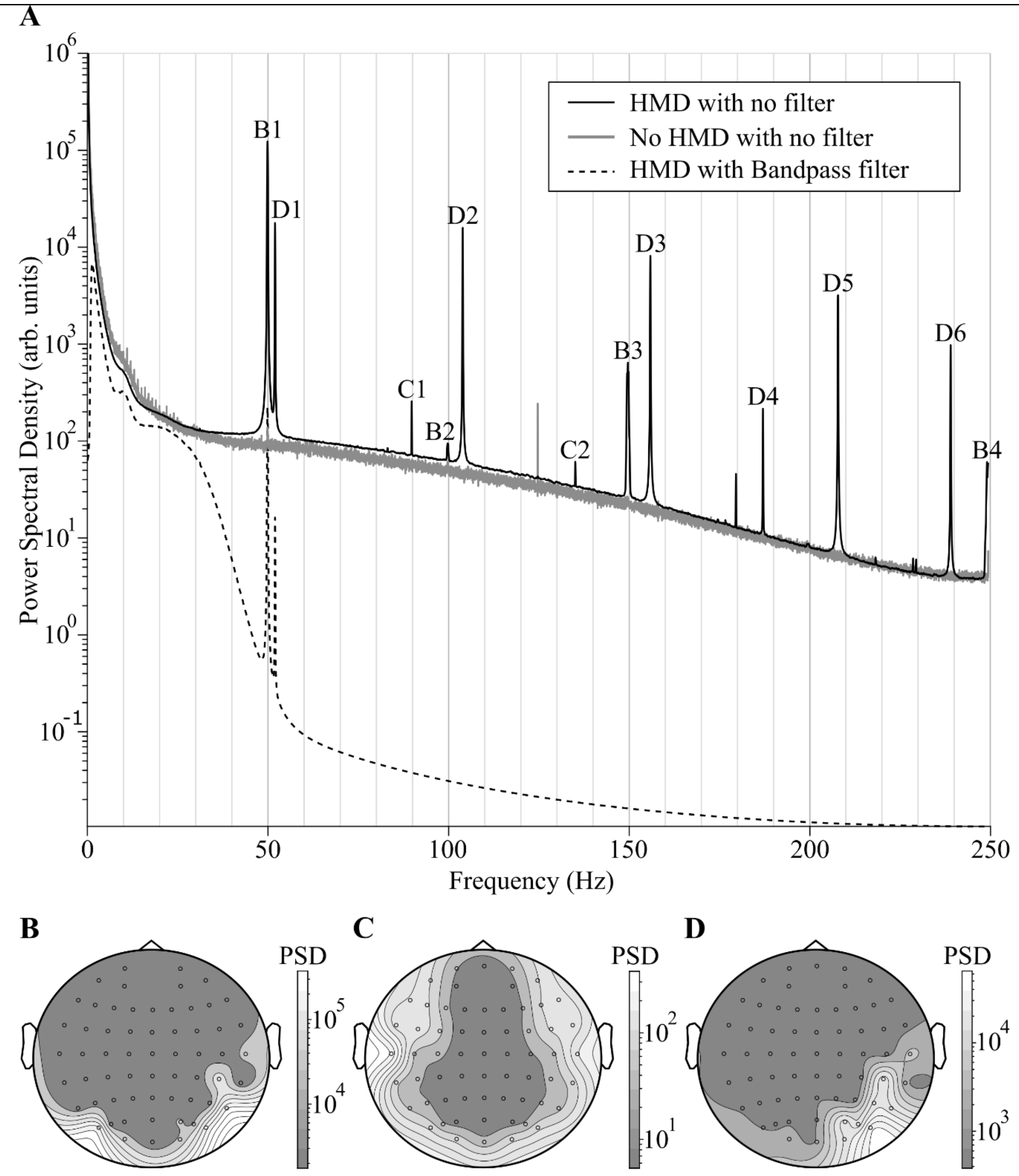


Figure 8. Comparison of the PSD of EEG recordings when a head-mounted display (HMD) is worn and when it is not. Distinct, high frequency peaks related to the HMD can be seen. The EEG recordings with the HMD on come from 54 participants who took part in one of our experiments. EEG data where no HMD was worn were acquired from an unrelated experiment ran by a colleague. These data were selected as they used the same EEG equipment. A) Shows the mean PSD of 62 EEG channels, averaged across datasets. In addition to the PSD of raw, unfiltered EEG from both datasets, the effect of a 1-35Hz bandpass filter is shown for the HMD on dataset. The frequency range considered is from 0-250Hz, as 250Hz is the upper limit observable at the 500Hz sampling rate, as defined by the Nyquist-Shannon sampling theorem.

Overall, this artefact, and others produced by the head-mounted display, were well above frequencies of interest and could be removed using standard filtering techniques, such as a bandpass filter. For researchers wanting to investigate higher frequencies, such as gamma (>30Hz), methods for removing line noise could be adapted (e.g. ZapLine; Cheveigné, 2019).

3.8. EEG analysis techniques

3.8.1. Data cleaning

In parallel with developing methods for neuro-VR and making considerations to allow EEG data to be recorded that was amenable to analysis, we needed to develop a suitable analysis pipeline. Throughout all experiments the analysis pipeline was consistent. Only small adaptations to specific datasets were required. These mostly related to spectral filters and the epoching of trials and will be described in the data preparations sections for each experiment.

After filters and epoching, we used a ‘pre-cleaning’ approach to aid the final cleaning of the EEG data. To do this, first an independent components analysis using the infomax algorithm (Bell & Sejnowski, 1995) implemented in Fieldtrip was used to identify sources of activity at the sensor level. The first ten components were reviewed and, if present, eye blink, eye movement and cardiac artefacts were removed. Review of components was assisted by a graphical user interface we implemented in MatLab which provided plots of component weighting across sensors and time, as well as the PSD of the component. This ‘pre-cleaning’ allowed for easier disambiguation of artefacts from EEG when the time course was event related. For example, it made it easier to distinguish some low frequency artefacts if they occurred on the same channel or around the same time as an eye blink artefact or a response involving movement. Every channel and trial of the ‘pre-cleaned’ data was visually inspected, and any artefacts were labelled. Trials containing one or more artefacts were noted. This list of trials was then used to remove trials from the original, non- ‘pre-cleaned’ dataset. A second independent components analysis was then run on the new list of trials. This time, all independent components were reviewed, using the same method as before. As before, eye blink, eye movement and cardiac components were removed, if present and clearly independent of brain activity. If needed to avoid excluding many trials or channels, additional sources of noise, such as muscle artefacts were removed as well. A final inspection of all channels and trials was then made. If artefacts were still present in some trials these were noted and the whole cycle of 'pre-cleaning' began again.

3.8.2. Statistical analysis

Comparisons between conditions were made using cluster-based non-parametric permutation test methods implemented in Fieldtrip (Maris & Oostenveld, 2007; Oostenveld et al., 2011). For brevity, I will refer to these tests as "cluster-based analysis". Like many inferential statistics, cluster-based analysis can be used to test the hypothesis of whether two sets of data belong to the same distribution. This test was chosen because it corrects for multiple comparisons without reducing statistical power as much as other methods, such as Bonferroni correction. In addition, it allows for biologically plausible constraints to be applied to the data. Full details are available in the cited paper, so I will only explain them in brief and in relation to the way they were used in our analysis.

The test relies on two key methods. One is the non-parametric statistical test between two datasets. These might be trials from different conditions or groups, for example. First, the datasets are appended to form a master dataset. The master dataset is then randomly partitioned to provide two datasets of equal size to the original ones. A suitable test statistic, such as a *t*-test, is then calculated between the random partitions. The results of this test (e.g. the *t*-value) are recorded and then another random partition is tested. This process should be repeated many times, as will be explained. The test statistic is also calculated between the original datasets. The ratio of the total number of partitions to the number of random partitions that had a higher test statistic than the original dataset should then be calculated. This figure is called the Monte Carlo estimate and determines the probability that the two datasets belong to the same distribution. If the probability is less than alpha (commonly .05) then the null hypothesis can be rejected. To provide sufficient accuracy, the Monte Carlo estimate needs to be based on a sufficiently large number of estimates. If, for instance, 1000 permutations are calculated then the estimate can be made to .001 accuracy. Precision of the estimate can then be measured using confidence intervals for p (Buckland, 1984; Ernst, 2004). When reporting cluster analyses, I will provide details of the 95% confidence interval (95% CI [lower upper]) for all p-values.

The second key method relies on forming clusters across one or more dimensions of the data structures. For sensor level time-frequency data, there are temporal, spectral and spatial dimensions. Some constraints need to be applied to these to ensure that clusters are only formed in biologically plausible ways and are related to the hypothesis being tested. For example, we can assume that data that are close in time, frequency and/or space are more likely to share a common signal source, so we can define sensible neighbourhood structures based on these assumptions. For time and frequency, neighbourhood is defined as adjacent samples. Note, the sample rate must be unbroken and high enough to justify the assumption that they share a signal. A spatial cluster is formed between neighbouring sensors. There are several plausible methods for defining sensor neighbourhoods. Common practice is to create a mesh of polygons from a 2D representation of the EEG cap layout and assign neighbourhood based on the vertices between electrodes (Maris & Oostenveld, 2007). However, we did not use an equidistant cap layout which meant that some parts of the scalp have a higher density of electrodes (Jasper, 1958). These would be underrepresented using this method. Therefore, we decided to use the template 3D electrode positions created from the average of all participants' head scans to determine neighbourhood. Electrodes were considered neighbours if they were 5cm or less from each other. This distance ensured every electrode was at least neighbours with adjacent electrodes.

These two methods can be combined as follows. Create a master dataset from the two datasets to be compared. They must be of equal size. Calculate a test statistic between every paired sample in the two new datasets. Label all comparisons with a test statistic above some threshold (e.g. the *t*-value determined to be significant based on the degrees of freedom of the sample). Form clusters of neighbouring samples if they are above threshold. A minimum cluster size can be set for each dimension. For example, a cluster may be required to be present in two or more channels. The test statistics within each cluster should be summed and the largest recorded. There are some differences between how this is done for one- and two-tailed tests and for positive and negative clusters which are discussed in the original paper (Maris & Oostenveld, 2007). Just as for single sample datasets, many permutations allow for a distribution of test statistics to be formed. If present, the summed test statistics of clusters found from the original comparison can be compared to the distribution to form the Monte Carlo estimate and determine significance.

3.8.3. Source analysis

3.8.3.1. Why source analysis?

We wanted to identify the source/s of EEG recorded at the scalp to aid our interpretation of the activity we had measured. One of the main limitations of sensor level analysis we used is that only the average effect of many sources can be observed. Further, the cluster-based analysis means that it is likely only one significant cluster can be found per analysis. This is because only the largest cluster is used to form the probability distribution that the Monte Carlo estimate is calculated from (Maris & Oostenveld, 2007). This can lead to problems interpreting clusters that form only the positive or negative component of a dipolar effect. For this reason, interpretation of sensor level clusters needs to be done with caution.

As we did not directly measure any independent sources in the brain, we had to apply an inverse solution that could estimate the most likely sources that could result in the EEG we recorded. Without applying constraints, there would be infinite possible solutions and they would not be informative about the real sources of activity in the brain. Fortunately, many methods exist to constrain the number of possible solutions to potential sources of activity in the brain. Collectively, these constraints are called the forward model.

3.8.3.2. Requirements for forward model

As the forward model describes how real sources are represented in the data collected, most methods for calculating it require two key pieces of information. First, the possible real sources, which in the case of EEG can be identified from a structural MRI. Second, how the data that was collected relates to those real sources. There are many possible methods for defining those two pieces of information, but for EEG knowledge of the electrode positions on the scalp in the same coordinate space as the structural MRI is commonly required. The structural MRI can then be segmented according to tissue type (scalp, skull and brain). The process will be described later, but with that information it is possible to create the forward model. For many EEG studies, including our own, it is not possible to collect structural MRIs for all participants. Further, equipment used for measuring electrode positions, such as a Polhemus digitiser (Colchester, VT, USA), is expensive and not available in many EEG labs. In our case, some data was collected remotely and our AFO participants were only able to commit a limited amount of time to the experiment. This meant that we had to have a way to create the resources we needed for source analysis that used mobile equipment and did not require much time from participants. As these are common requirements in EEG studies there are several methods that can be used to meet them.

For electrode positions, these may be provided by EEG cap manufacturers or within analysis tools. These templates are not accurate when compared to individual measurements (Homölle & Oostenveld, 2019) and cannot account for the great variation in head size and shape. It is not common practice to account for superior inion-nasion or left-right pre-auricular distances. Further, templates from manufacturers are based on their offered cap sizes which typically have broad fit allowances, such as 57-61cm. Assuming accurate measurement, this means the template will on average be 2cm in error (Homölle & Oostenveld, 2019). Inaccurate electrode positions have been shown to introduce error into the forward model for EEG (Dalal et al., 2014), negatively impacting source analysis. Therefore, we sought an alternative method of measuring electrode positions. The most suitable alternative was to use a structured light scanner to produce a 3D model of the head. There are several formats for these devices, including handheld devices which you move around a stationary object. They can be used to quickly produce accurate reconstructions not just of the shape but colour as well (Rocchini et al., 2001). This speed is useful for testing participants who cannot remain still for extended periods of time (Ettl et al., 2013). In our case we found it useful when collecting data from participants with time constraints, such as police officers. Like head-mounted displays, the development of structure light scanners has been propelled by commercial application in industries such as architectural and graphic design. This means that, relative to specialist scientific equipment, such as Polhemus digitisers, they are also cheap and able to interface with many other devices and software.

3.8.3.3. Creating the forward model

The first step for creating the forward model was to define the possible positions of sources in 3D space and their relation to all the electrodes that may have recorded their signal. This could be done using the tissue boundaries for scalp, outer skull and inner skull, and electrode positions attained using the structure light 3D scan methodology we developed (Figure 9). However, it was important that the forward model for each participant described the same number of potential sources and that they corresponded to approximately the same anatomy. This was so that contrasts could be made for activity in source space between participants. Otherwise, contrasts would have only been able to be made after interpolating individual source solutions to a common space, which requires a great deal more computation time. Therefore, we first defined positions based on the tissue boundaries of the ICBM152 template (Collins et al., 1999; Fonov et al., 2009), from which all individual MRIs and boundaries were made from, and then matched individuals' forward models to it.

The head model of the template was created using the Boundary Element Method (Oostendorp et al., 1989). BEM fits the outermost boundary (scalp) into a cubic grid. For the template this grid was 18.5cm by 14.5cm by 15.5cm volume. We decided to use $5mm^3$ voxels as an appropriate balance between computational time and high resolution needed to appropriately define tissue boundaries. The tissue boundaries for scalp, outer skull and inner skull were used to define each voxel as brain, skull or scalp tissue using the 'bemcp' function implemented in Fieldtrip (Phillips et al., 2000). Any voxels not defined were marked as outside the skull. The same process was applied to individual tissue boundaries, using the same $5mm^3$ voxel size. However, because the outermost boundary was not the same size as the template's, the dimensions of individual grids varied in size. The function 'spm_normalise' in SPM12 was used to apply a non-linear spatial normalisation (described in Ashburner & Friston, 1999) of individual grids to the template grid, such that for any voxel coordinate, every participant's grid would have the same tissue label. These voxels would also correspond to approximately the same anatomy.

All voxels were assigned values based on the conductivity of their assigned tissue (0.33, 0.0042 and 0.33 Siemens per meter for scalp, skull and brain, respectively; Gabriel et al., 1996). We then identified voxels within the cortex using the MatLab function 'inpolyhedron' (Sven, 2019) to determine whether voxels were inside a mesh of the left and right cortices defined by the Mindboggle Atlas (A. Klein et al., 2017). Using the electrode positions and conductivity values, a lead field, which estimated how a standard current (1 Amp) at each voxel would be recorded at each electrode, was created using Fieldtrip (Oostenveld et al., 2011; originally described by Scherg & Von Cramon, 1986). For a more detailed example of this procedure, see Oostenveld and Oostendorp (2002). With these estimations of the relationship between what we could and could not observe, we were now able to find more valid source solutions for the EEG recorded during the experiments.

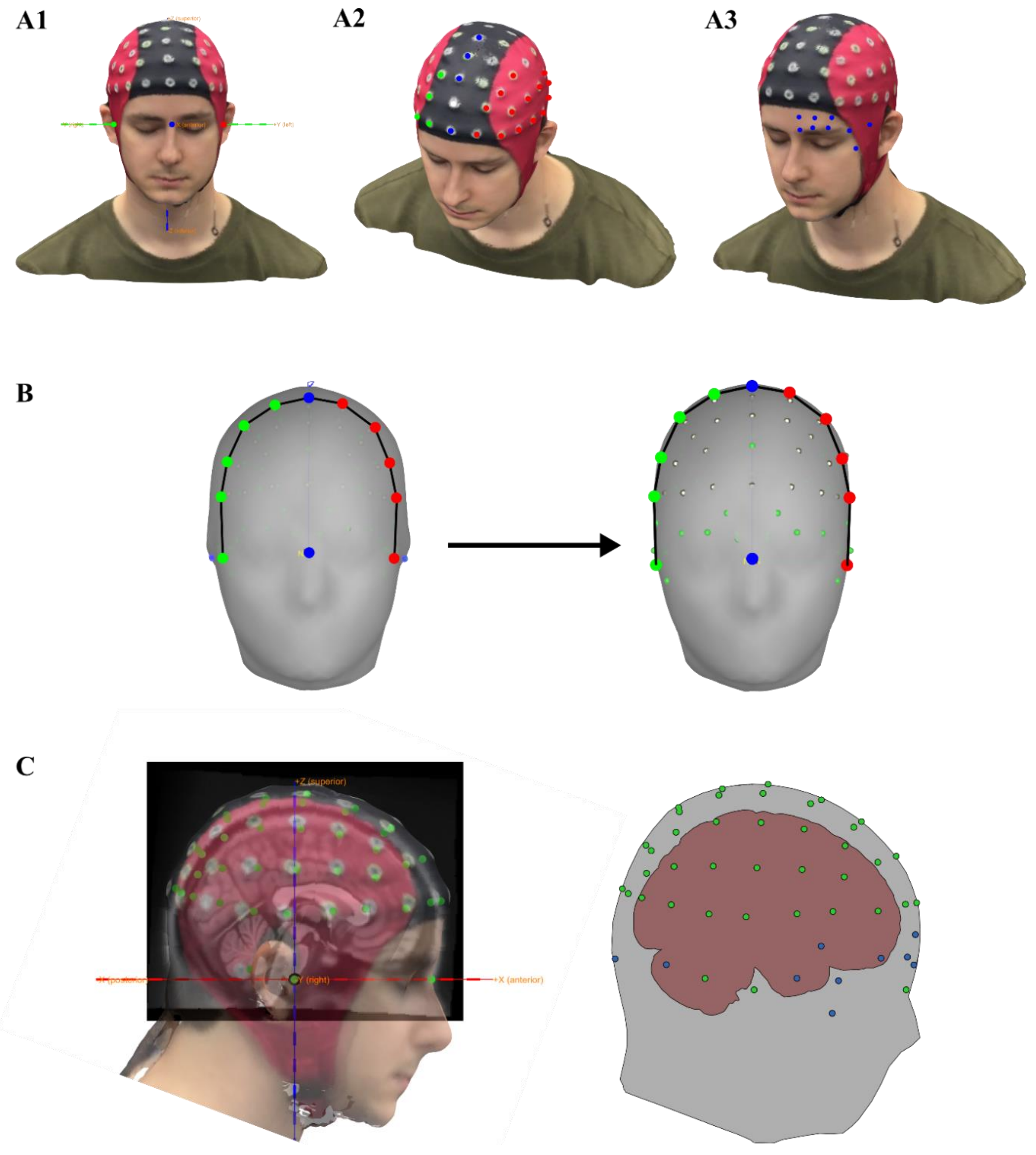


Figure 9. Method for warping structural MRI and BEM to fit individual 3D scans. The 3D scan used here is of the author. A) The structured light scan was imported into MatLab. Fiducials, electrode positions and additional points of the skull were labelled in Fieldtrip (Homölle & Oostenveld, 2019; Oostenveld et al., 2011). Electrode positions were moved in towards the centre of the head by 8mm, as that was the thickness of the electrode. B) Coordinates of the marked points were passed to Brainstorm (Tadel et al., 2019) where they were treated as measurements of head shape. A template anatomy (ICBM152; Fonov et al., 2009) was then warped to fit these landmarks using a nonlinear transformation (Ermer, 2001). C) The warped structural MRI and BEM was passed back to Fieldtrip and visually inspected for errors.

3.8.3.4. Source localisation using eLORETA

To calculate the inverse solution, we used the Fieldtrip implementation of eLORETA (exact low-resolution brain electromagnetic tomography; described in Pascual-Marqui et al., 1994) eLORETA can estimate the current density in cortex in three-dimensions, prioritising localisation over spatial resolution. The assumption is that neighbouring voxels have highly correlated current density. We found it particularly well suited for our data, as estimating sources with high spatial resolution would rely too much on our limited number of electrode channels. eLORETA has been validated several times using similar EEG datasets (e.g. Jatoi et al., 2014; Smith et al., 2019). While it has been shown that performance of eLORETA is comparable to other inverse solutions, such as minimum-norm estimation and beamformer techniques (Halder et al., 2019), there can be considerable variation in outcome, depending on the exact implementation used (Mahjoory et al., 2017). We decided to use eLORETA because of its suitability for our dataset. Specifically, our EEG data had a moderate number of electrodes, and could have artefacts removed during pre-processing (eLORETA does not explicitly filter noise). Our forward model was also tailored towards use by eLORETA, as we constrained sources to cortex. Because of this, comparisons of results with other inverse solutions were outside of the scope of the project.

3.8.3.4. Source statistics

Cluster-based analysis can be applied to source localised EEG data. From the point of view of the test, the input and output are essentially the same. The main difference is the increase in the number of data points compared. At sensor level, we formed clusters from 63 data points on the scalp, which is small compared to the 14,075 defined inside the cortex. While the difference seems large and increases the multiple comparison problem greatly, the corrections made by the non-parametric tests overcome them in the same way (Maris & Oostenveld, 2007). This makes sense as they are only adapted from tests designed for analysis of voxels from fMRI data (Holmes et al., 1996).

With greater numbers of comparisons, it is much more likely that a comparison between two data points will result in a test statistic above the threshold. For this reason, the reliance on cluster formation is emphasised. As will be seen in the reporting of results, many small clusters are likely to be found because of the reduced possibility of overlap, relative to sensor level analysis. Evaluation of the

Monte Carlo estimation allows these spurious clusters to be dismissed in favour of larger, dominant clusters. When reporting cluster statistics, I will state the total number of clusters and then describe only significant ones.

A few checks can be made that the inverse solution is an acceptable approximation of the real sources. This can be done by comparing these estimations to the real measures at the scalp level. In general, the direction of effect should be matched between comparisons at source and sensor level. However, the location of significant sources may not match the sensor level analysis; a positive frontal cluster at sensor level does not necessarily mean the source is a frontal region of the brain (Van de Steen et al., 2019).

3.9. Summary of neuro-VR methods

After considering many alternative virtual reality methods to combine with EEG, we decided to use an Oculus Rift CV1 head-mounted display. We created a development pipeline for creating experiments to present in virtual reality based off technology used in the gaming industry. This development pipeline was centred around UE4. With it, we could present scenarios involving animated virtual humans with natural movements within a realistic 3D environment. Crucially, these scenarios would meet the demands of our experimental design, by recording precise timings and log files as a record of events within them.

Alongside development of the virtual reality side of ‘neuro-VR’, we adjusted more established pipelines for EEG analysis. This included taking additional pre-processing steps to remove artefacts introduced by natural movements and the proximity of electrical components of the head-mounted display. Further, we ensured that data could be collected using a mobile lab without loss of information such as individual structural MRIs and electrode positions. With this data, we would be able to use source analysis techniques to test our hypotheses.

Chapter 4: Contextual Prop and Action Experiments

4.1 Introduction

The methods we developed for neuro-VR, described in the previous chapter, meant that we could have begun developing complex AFO training scenarios into an experimental format. However, before we did this, it was important that we tested our new methods with a simple paradigm. In this sense, one aim of the empirical work presented in this chapter was to further validate our neuro-VR methods. A second aim was to investigate the behavioural and electrophysiological differences between the decision to shoot or not in police decision making. By doing so we would have a basis to work on in our later research.

With these aims in mind, we developed two experiments made up of simple, repeated 'shoot/don't-shoot' scenarios. These scenarios were developed with guidance from expert police firearms instructors and AFOs. We discussed with them the kinds of situations AFOs are tasked with resolving and observed the training they undergo to prepare for them. From these discussions we identified two manipulations that were present in most training scenarios and grounded in real world incidents: threat level and compliance. Threat level was often determined by whether a weapon was present, and the level of threat it presented. Compliance related to how responsive a perpetrator was to instructions from an AFO. These manipulations formed the basis of the two experiments presented in this chapter. The first being the Contextual Prop experiment, in which participants decided whether to shoot a virtual human based on what he was holding in his hand when raising it towards them. The two props used were a handgun and a can of drink. The second experiment, called the Action experiment, varied the behaviour of a virtual human holding a handgun so that they were either compliant (surrendering) or not (attacking).

Several studies have used similar manipulations as ours in 'shoot/don't-shoot' experiments. Johnson et al. (2018) compared data collected from novices and police officers in a series of experiments where they manipulated prior information given to participants about a perpetrator. One of the manipulations of the perpetrator in the scenarios was whether they were armed with a handgun. Usefully, they measured reaction time for both 'shoot' and 'don't-shoot' responses. They found that in all experiments participants in both groups were faster to respond to armed versus unarmed perpetrators. This is consistent with other ‘shoot/don’t-shoot’ studies where reaction times for both conditions have been measured (Correll et al., 2002, 2006; Fleming et al., 2010). It is more common for ‘don’t-shoot’ responses not to be measured (e.g. Johnson et al., 2014; Petras et al., 2016) in keeping with traditional Go/No-Go experiments where inaction is critical to the paradigm (Benikos et al., 2013). In the case of ‘shoot/don’t-shoot’ experiments, whether or not ‘don’t-shoot’ can be measured depends on the task setup and instructions. For a button pressing task, the two responses are comparable, but when accurate shooting is required additional consideration needs to be made for the time it takes to aim. Nieuwenhuys et al. (2012) navigated this issue by asking participants to shoot at a black square proximate to the perpetrator for their equivalent of a ‘don’t-shoot’ condition. Perhaps due to the requirement to aim in both conditions, they found a larger difference between conditions than is typically reported. Based on this research, we predicted that response times would be greater in non-threatening scenarios. Further, when designing our experiments, we ensured that reaction time measurements would be comparable across conditions. This will be discussed in the methods section.

In the ‘shoot/don’t-shoot’ studies I have mentioned, the reason for faster responses to threatening stimuli is not discussed. Most likely this is because this main effect is usually dominated by interaction with other variables of more interest to the study (e.g. prior information or stress; Johnson et al., 2018; Nieuwenhuys et al., 2012). Faster reactions in response to threat can be attributed to preferential, adaptive orientation towards threatening stimuli (Lang et al., 1990; Öhman & Dimberg, 1978; Seligman, 1970). Participants in ‘shoot/don’t-shoot’ experiments are in a state of alertness as they are expecting to have to respond to a stimulus (Posner & Petersen, 1990). This is emphasised by the possibility of threat (Öhman et al., 2001). Further, alertness may accelerate the effects of orientation (Callejas et al., 2004) –in this case towards the threatening stimuli. It may be that if participants were not alert to the perpetrator then no difference in reaction time or related performance would be observed, but this has not been tested to date. The cognitive functions that maintain alertness and allow for selective orientation towards aversive stimuli are closely related to the basal ganglia structures of the reward circuit and in particular the ventral striatum (Graybiel & Grafton, 2015; Jin & Costa, 2015; Saga et al., 2017). Through these pathways, increased attention and motivation should lead to the faster reaction times observed in ‘shoot/don’t-shoot’ experiments. Differences in activity at the basal ganglia and their connections, in regions such as the anterior cingulate cortex (ACC) and pre-frontal cortex (PFC), are likely candidates for differences in EEG signal during our ‘shoot/don’t-shoot’ tasks (Khalighinejad et al., 2020).

Our project was not the first to use EEG to investigate ‘shoot/don’t-shoot’ decision making. Correll et al. (2006) examined ERPs to investigate the effects of stereotypes on behaviour in a ‘shoot/don’t-shoot’ videogame-based task. They reported similar stimulus-locked ERP waveforms, in response to both armed and unarmed perpetrators. The main significant differences between conditions were greater amplitude N100 and N200 components, although later components differed as well. Here, the N200 component is of interest as it replicates findings from Go/No-Go tasks, from which we know it is associated with greater response inhibition (Nieuwenhuis et al., 2003). It could be argued that to shoot requires inhibiting pressing the ‘don’t-shoot’ button in this case, but within the broader context of the task there is a clear prepotent response to shoot. The source of the N200 signal is thought to be the ACC (Bekker et al., 2005) and its amplitude is positively related to frontal-midline theta power (Hajihosseini & Holroyd, 2013). Frontal-midline theta has been related to alternating activity between the ACC and PFC (Asada et al., 1999). It may be related to conflict monitoring or cognitive control and may act as the mechanism by which control is realised (Botvinick et al., 2004; Cavanagh & Frank, 2014). Because of this, we expected to find early differences in frontal-midline theta between threatening and non-threatening scenarios in our experiments.

Brain activity related to the ACC has also been the target of investigations into errors in shooting tasks. Fleming et al. (2010) recorded EEG from military cadets completing a ‘shoot/don’t-shoot’ task and compared error-related negativity (ERN) between conditions. ERN is a negative, early (typically sub-150ms), response-locked ERP, related to decision errors (Holroyd & Coles, 2002), and localisable to the ACC (Herrmann et al., 2004). Note, ERN can be detected in correct trials and may reflect the same processing in the ACC as the N200 (Holroyd, 2004) and is also related to frontal-midline theta (Luu et al., 2004). As expected, Fleming et al. (2010) found that the amplitude of ERNs was greater for false-positive (shooting in non-threatening trials) but not false-negative (not shooting in threatening trials) errors. One limitation of their study was that participants made few errors in some conditions. Generally, participants in ‘shoot/don’t-shoot’ tasks make very few mistakes (Landman et al., 2016b;

Nieuwenhuys, Oudejans, et al., 2012), which may be due to the relatively unconstrained reaction time (Wood & Jennings, 1976) thresholds used and the emphasis on making the right decision. This is a positive result, especially when the population being studied is an expert one, as it represents the desired state of 'shoot/don't-shoot' actions in the real world. However, in research, some errors are needed in all conditions for appropriate statistical analyses to be used. Given our intention to analyse response-locked EEG, post-response theta related to error-processing in the ACC was a target of interest in our analysis, but this was conditional on participants making enough errors.

We did not plan on contrasting data from the Contextual Prop and Action experiments directly, in favour of treating them as independent investigations. Primarily this was because we were more interested in the commonalities between the experiments than differences. It was possible that participants could treat the tasks as fundamentally different, where one relied on object discrimination and the other on action perception. However, we expected that the task instructions and the emphasis on response would be more important. Ultimately, our predictions about the commonalities and differences between the experiments needed to be quite tentative. Given the task's simplicity, they were unlikely to be sensitive to subtle differences in perception. For this reason, outside of our main comparisons between threatening and non-threatening scenarios, much of our analysis was exploratory. Findings from these analyses would guide our planned research comparing expert AFOs against novices in a similar task. In a sense, the Action and Contextual Prop experiments were designed to form the groundwork for that investigation. Critically, findings about the effects observable in 'shoot/don't-shoot' paradigms would allow us to make predictions about how they would be moderated by expertise.

4.2 Method

4.2.1. Research design

Both experiments had the same single factor, two level design and only varied in how the independent variable manipulated threat. These are described by the names of the experiments: either the Action (Attack vs. Surrender) or the Contextual Prop (Handgun vs. Can) determined whether a trial was threatening. In both, we measured reaction times and the validity of participants' responses as a measure of performance. This was alongside our primary measure, EEG, which we used to measure electrophysiological changes in the brain. Specifically, we measured change in power, relative to baseline, in neural oscillatory frequency bands around the presentation of decision-relevant stimuli and participants' responses to those stimuli.

4.2.2. Participants

A total of 32 participants completed both experiments in one approximately two-hour long session. This sample size was not based on any power calculation but was rather the result of data collection over a pre-determined period. A power calculation would not have been appropriate as our hypotheses were varied and included planned exploratory analysis. Further, to our knowledge, no similar paradigms have used comparable measures to our own, so we were unable to meet the required assumptions for a power analysis (Bacchetti et al., 2011). A single participant was removed from all analyses due to a failure to follow task instructions. The order of experiments was counterbalanced for age and sex. This was to avoid order effects related to improvements in performance and/or engagement with the task. Participants were recruited from either the general population or were university students or staff. Prior to participating in the study, we screened all participants for the following exclusion criteria, assessed by self-report: unable to stand comfortably for between one and two hours; diagnosis of photosensitive epilepsy or traumatic brain injury; sensitivity to motion sickness or cybersickness. All participants had normal or corrected to normal vision. Where possible, participants were asked to wear contact lenses rather than glasses to increase comfort when wearing the head-mounted display. All participants gave their informed consent to participate in the experiment. This research was given a favourable opinion by a university research ethics committee (ref: 1179).

4.2.3. Materials and apparatus

4.2.3.1. Physical setup

During the experiments, participants stood in the centre of an area which they were instructed not to move from. A protection system built into the head-mounted display software, called Guardian (Valve Corporation, Bellevue, WA, USA), warned them if they approached the edges of the area (approximately 4x4m) by presenting a blue wall at the perimeter. They were also continuously monitored by the experimenter. The only object in the area was a surface for putting the EEG equipment on. There were also cables for the EEG and head-mounted display equipment, which participants were made aware of.

4.2.3.2. Virtual setup

In both experiments, trials were presented within a realistic-looking 3D virtual environment. To minimise distractions and ensure good graphical performance, the environments were kept simple: a small grass courtyard surrounded by brick walls with a natural blue sky. It was created using the development pipeline described in the previous chapter.

We decided to only use one virtual human for all trials in both experiments. This was because we did not want to introduce unnecessary confound variables into the experimental designs. If we were to use multiple virtual humans, their number would have to match the total number of trials in the experiment. Otherwise, participants may have considered the outcome of their last interaction with a virtual human when they appear in new trial. This may have been particularly salient for trials in which a participant made a mistake and was shot. Participants may have been made more biased to shoot them the next time they appeared. Producing one virtual human per trial would have been highly time consuming, and may have introduced other biases related to, for example, ethnicity, which were outside the scope of our investigation. Therefore, we decided to go with the simple option of using just one virtual human, a white male.

4.2.3.3. EEG measurement

Participants wore a 65-electrode (Ag/AgC) ANT Neuro (Hengelo, Netherlands) waveguard EEG cap underneath the head-mounted display, and the placement of electrodes within the cap followed International 10-20 convention (Jasper, 1958). Conductive gel was applied to each electrode to ensure good conductivity with the scalp. This setup was non-invasive and took approximately 20 minutes. During on-line recording, electrodes were referenced to the CPz electrode and grounded at position AFz. All electrodes were continuously sampled at 500Hz.

Before putting the head-mounted display on, a 3D scan was taken of the participant’s head, from the shoulders up. This was done using a structured light sensor (Occipital Inc., Boulder, CO, USA) for the purpose of creating individual forward models using our established pipeline. Only one scan was taken per participant, with the assumption that the cap did not move between the two consecutive experiments.

4.2.4. Procedure

4.2.4.1. General procedure

Participants were informed that they would complete two experiments which contained repeated scenarios of the kind AFOs respond to. They were instructed to resolve each scenario by either shooting the virtual human or indicating that they were not a threat. To shoot a virtual human they would aim a virtual gun using their dominant hand and pull a trigger with their index finger. They were told that they could miss if they did not aim at the virtual human. If they hit, the virtual human would fall to the ground. To indicate that the virtual human was not a threat, they were told to press the Safety button with the thumb of their non-dominant hand. At this point they would hear a click which meant they could not change their mind and discharge their firearm. More obviously, it was also emphasised that once they shot the virtual human, they could not then decide that they were not a threat. We asked that responses be as quick and accurate as possible, regardless of the action.

Participants were also given some instructions about how they could help ensure the EEG was as artefact free as possible. Specifically, they were asked not to talk, unless to the experimenter, and to try and keep as neutral an expression as possible. In addition, they were asked to face forwards, without looking around the environment, while continuously aiming the firearm at the virtual human during trials. To reduce muscle artefacts in the EEG data caused by continuous aiming, they were asked to aim simply by putting their arms at their side and bending their elbow to point their firearm at the virtual human. Another reason we gave this instruction was to ensure that recorded reaction time more accurately reflected the time at which participants decided to shoot by removing variability involved with movement of the arm. While this would remove aiming, which is an important aspect of performance, we had already decided to make the target very easy to hit and so the task was not sensitive to this measure.

After this general instruction, participants were given more specific instructions about the first experiment they would be completing. They then completed a practice task, followed by the experiment itself. The second experiment was then explained, practiced, and completed. At each stage we checked for symptoms of motion sickness. Breaks to sit down and drink some water were also offered between stages.

Each experiment lasted approximately twelve minutes and comprised eight blocks of ten trials, where each trial included one scenario. Between each block, participants could rest and sit down, but none opted to remove the head-mounted display. Trial conditions were randomised within blocks, so that each contained an equal number of threatening and non-threatening trials. Trial structure was matched between the two experiments, but the stimuli and some timings varied, as described below.

4.2.4.2 Action experiment trial procedure

For the Action experiment, all trials began with the virtual human appearing 2m in front of the participant, standing in a neutral stance with hands at his side. He could be seen holding a gun in his right hand. After 3s he would either raise his right hand and point the gun at the participant or raise both hands in a surrendering motion. These were the threatening and non-threatening actions, respectively. Both animations took 1s to complete, after which the virtual human would hold his position for a further 2s.

4.2.4.3 Contextual Prop experiment trial procedure

In the Contextual Prop experiment, all trials began with the virtual human appearing 2m in front of the participant, but standing behind a low wall, the height of which was automatically calibrated against the height of the participant so that they would not see what was held in the virtual humans' hand. In all trials, the virtual human would raise his right hand after 3s and hold it up towards the participant. In his hand would be either a blue can of drink or a handgun. In both instances the animation of the virtual human was identical, but the prop gave it context: either he was holding a can out, non-threateningly, or he was pointing a handgun, threateningly. The effect of obscuring the prop behind the wall was that it was only completely revealed to the participant 400ms after the start of the animation.

4.3. Results

4.3.1 Behavioural analysis

4.3.1.1 Data preparation

Data from both experiments were treated in the same way. MatLab (The Mathworks, Inc, Natick, MA, USA) was used for all stages of data preparation. First, the log files produced for each trial were imported and collated, after which they were assigned to the associated participant number and demographic information on age, sex, and handedness.

Performance at the tasks was assessed independently, as, despite the similarity of the two experiments, it was possible that one task was misunderstood and the other was not. Participants were removed from the dataset for an experiment if the percent of correct responses in one or more conditions was close to chance. This threshold was set for two reasons: first, poor performance in any one condition indicates misunderstanding, lack of attention, or unexpected difficulty with the task (pilot testing of the paradigms suggested performance should be near ceiling). Secondly, pre-processing and subsequent analysis of EEG data requires many repetitions of each trial to ensure good signal to noise ratio. Removal of too many trials due to incorrect responses would therefore reduce the signal to noise ratio of the EEG data from that participant. Therefore, any participants removed from behavioural analysis were also removed, in total, from the EEG analysis.

Two participants were removed from the Action experiment and five were removed from the Contextual Prop experiment. Their percent correct ranged from 0% to 62.5%, relative to the mean percent correct of 94% and 95% in the Action and Contextual Prop experiments, respectively. Interestingly, there was no overlap in the participants removed from each experiment and, further, there was no clear relationship between the order of experiments and the experiment they performed poorly in. Descriptions of the remaining participants can be found in Table 1.

Table 1. Description of participants included in the behavioural data and EEG analysis for the Action and Contextual Prop experiments. Also shown is the effect of removal of participants on the counterbalancing of experiment order for age and sex.

Experiment Order	Included in Action Experiment		Included in Contextual Prop Experiment	
	Age (*SD*)	n (men/women)	Age (*SD*)	n (men/women)
Action → Contextual Prop	27.8 (4.7)	14 (7/7)	26.8 (4)	12 (4/8)
Contextual Prop → Action	26.2 (7.7)	16 (7/9)	26.9 (8.1)	15 (6/9)
Total	26.9 (6.4)	30 (14/16)	26.9 (6.5)	27 (10/17)

Having removed these participants, outlier trials were identified and removed. This was done more to benefit the EEG analysis than the behavioural analysis. As will be described later, analysis of response-locked EEG requires non-overlapping epochs. Because of this, when designing the experiment time was allowed between each trial to allow for variability in response time. However, if participants responded slower than anticipated then the end of the response-locked epoch would overlap with the beginning of the next. Therefore, we needed a way of reliably detecting outlier trials and removing them from analysis. Without this need, a median average would have been used to similar effect, without needing to remove any data. The upper limit of reaction time was determined as three standard deviations above the mean individual response time. A lower limit for reaction time was also set. If the response time was less than 250ms, the trial was removed, as it was likely the predictable timing of the stimulus allowed any participant reacting that quickly to initiate their response before the stimulus was presented. See Table 2 for details about removal of outlier trials from the behavioural and EEG analysis.

Table 2. Number and reason for removal of trials. Note, trials in which the participant missed were not removed. Approximately 5% of trials were removed as outliers.

Experiment	Incorrect Response	No Response	Missed Shot	Outlier (lower limit)	Outlier (upper limit)
Action	77	6	31	3	31
Contextual Prop	95	3	33	0	25

Only correct responses were considered for analysis, and so trials in which participants' responses were incorrect, or they made no response at all, were removed. Incorrect trials were not analysed because, having excluded the participants who made the most mistakes as outliers, there were very few mistakes made across all participants. Either many more trials, or a more challenging version of the paradigm, would be needed to elicit enough incorrect responses to be analysed. The median average of response times within-subject was therefore taken from correct responses only.

4.3.1.2 Behavioural data results

The prepared data was exported from MatLab for statistical analysis using R: A Language and Environment for Statistical Computing (R Development Core Team, 2019). Just as for the data preparation, the data analysis was the same for both experiments. A dependent *t*-test was used to test whether there was a difference between the conditions of the experiments.

In the Contextual Prop experiment, response times to shoot the virtual human (M = 480.7ms, SD = 62.67ms) were significantly faster when compared with response times to press Safety (M = 540.2ms, SD = 76.05ms); $t(26) = 4.92$, $p < .001$, $d = 0.86$. The same effect was observed in the Action experiment, where shoot responses (M = 651ms, SD = 70.24ms) were significantly faster compared with press Safety (M = 730.5ms, SD = 131.42ms); $t(29) = 4.20$, $p < .001$, $d = 0.79$. In both instances, the effect size was large.

We conducted exploratory analysis of the effects of age and sex on response times. These analyses were of interest, because for our next study we planned on controlling for age and sex, between groups. A two (between-subject, sex: male vs. female) by two (within-subject, response: shoot vs. press Safety) mixed factor analysis of variance was used to test for a main effect of sex and/or interaction with the observed significant main effect of response on response time. In the Contextual Prop experiment, no significant main effect $F(1, 25) = 0.08$, $p = .77$, $\eta_{ges}^2 < .01$, or interaction, $F(1, 25) = 0.89$, $p = .35$, $\eta_{ges}^2 < .01$, was observed. As expected, given the overlap in participants across experiments, no main effect of sex was observed in the Action Experiment, $F(1, 28) = 0.15$, $p = .7$, $\eta_{ges}^2 < .01$, and neither was a significant interaction between sex and response found, $F(1, 28) = 0.07$, $p = .79$, $\eta_{ges}^2 < .01$. No significant correlations between age and response times in either experiment was found.

4.3.2 EEG analysis

4.3.2.1 Pre-processing of EEG

We used the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld et al., 2011) for performing all data pre-processing steps on the EEG data. Due to the similarities between the two experiments, these steps were shared for both sets of recordings. Initial visual inspection of the raw data revealed that the recording of electrodes M1 and M2 was of poor quality in many participants. We suspected this was due to the loose fitting of the EEG caps around the mastoid bones and so decided to remove them from all datasets. Recordings from the remaining channels appeared to be good quality. The median average impedance was 5.9 kΩ. Data was re-referenced from CPz to the common average of all electrodes, excluding M1 and M2. The following pre-processing steps were applied to recordings individually, but the process was kept as consistent as possible.

Before dividing the continuous EEG into epochs, a band-pass filter (2-35Hz passband, zero-phase, two-pass [forward and reverse], Hamming-windowed, fourth-order digital Butterworth filter, -24dB/octave slope) was applied. A low-pass of 35Hz was chosen based on the investigation of high-frequency artefacts inherent in our data, as described in the previous chapter. A high-pass of 2Hz allowed easier visual inspection of the data and helped reduce noise from low frequency artefacts (cable sway, skin artefacts) related to movement (Klug & Gramann, 2020). This meant that we could not later analyse EEG activity in the delta frequency band, but due to the short trial epochs, time-frequency analysis of frequencies lower than 2Hz would not have been appropriate anyway, as will be discussed later.

Filtered, continuous EEG data were segmented into non-overlapping epochs defined from 3.9s before stimulus presentation, and 1.2s after the participant's response in each trial. Epochs had to be non-overlapping to avoid correlation between trials. This was important because cluster-based analysis assumes statistical independence of data being compared (Maris & Oostenveld, 2007). Stimulus presentation was defined separately for the experiments: for the Action experiment, the stimulus was the onset of the animation and for the Contextual Prop experiment it was defined as the moment the prop could first be seen (400ms after the onset of animation). Trials identified in the behavioural analysis as outliers were automatically removed at this stage. In addition, only trials with correct responses were included in the EEG analysis, so incorrect or 'no action' trials were also removed. Trials containing artefacts were also removed using the 'pre-cleaning' approach described in the previous chapter (< 5%).

4.3.2.2 Time-frequency analysis

Full-length pre-processed trials, which varied in length depending on response time, were redefined to form two datasets: one stimulus-locked, ranging from 3.9s before stimulus onset to 2.65s after; the other response-locked, ranging from 4s before response to 2.5s after. Zero-padding was applied to both datasets so that they had equal lengths of 7s. Note, the irregular lengths of these epochs were determined by the maximum non-overlapping epoch allowed by variation in response time across all participants.

Time frequency analysis was conducted for frequencies 2-35Hz using Fieldtrip's multi-taper-method convolution. Unlike the name suggests, we used only a single Hanning taper for all frequencies. Power at each frequency was calculated within sliding time windows. The width of these time windows was set at three times the wavelength of the specified frequency. For example, at 5Hz the time window was 600ms wide. This is the reason that frequencies lower than 2Hz were not analysed. The central point of the time windows was moved in steps of 50ms. The effect of the sliding window and Hanning taper was that estimates of power across time and frequency, within each trial, could be made.

Baseline correction was then applied to the time-frequency data using the following method. First, the baseline period was defined for the stimulus-locked dataset from 3.75s to 3.25s before stimulus onset. During this period participants were standing while waiting for the virtual human to appear and the trial to begin. Within each frequency, power was averaged across the baseline period to produce a vector of power over frequency (*B*). Decibel conversion was then applied to the frequency (*f*) by time (*t*) activity matrixes (*A*) of the stimulus- and response-locked datasets, resulting in baseline corrected data (*C*) for each. The formula for this method of baseline correction is below:

$$C = 10 \times \log_{10} \frac{A_{ft}}{B_f}$$

4.3.2.3. Time-frequency results

4.3.2.3.1. 'Sanity checks'

A shared aim of the Contextual Prop and Action experiments was to further validate simultaneous use of head-mounted displays and EEG. In the previous chapter, investigations of the effect of head-mounted displays on EEG were presented. Generally, the conclusions were positive, with acknowledgement of limitations when analysing higher frequencies (>35Hz). However, it was still important to make some 'sanity checks' of data quality as this was our first attempt to analyse EEG data recorded during a virtual reality experiment. An overview of these checks can be seen in Figure 10.

We expected that around the response, frontal-midline theta activity would increase, and posterior alpha would decrease, relative to the baseline period. We also expected to see decreases in beta activity over the sensorimotor cortex. If the data was contaminated with artefacts, then these changes would be obscured and would not dominate the topographies. For example, if high-frequency artefacts from the virtual reality equipment or neck muscles were present then we would expect these to be clearly visible around the extraneous sensors (Figure 8, Chapter 3). We also wanted to check that our baseline period was suitable. This would be evidenced by little or no changes in power during or near the baseline period.

To make these checks, for each experiment we averaged the activity across both conditions. The mean average was calculated with equal weighting of conditions, regardless of the number of trials in each. This avoided over representing the condition with the least trials removed, but likely only had a marginal effect. The average was also made from baseline corrected data, rather than the raw data. We then inspected the response- and stimulus-locked time-frequency data. Note that for the three frequency bands of interest, positive and negative changes in power relative to baseline need to be interpreted in different ways. This is because the typical change in power during activity periods, relative to baseline, varies between them. Changes in theta band should be compared with a measure of relative increase in power versus baseline. In contrast, changes in alpha and beta should be compared with a measure of relative decrease in power versus baseline.

In Figure 10, panels (A) and (C) show that both experiments have very similar patterns of activity over time and frequency. Panels (B) and (D) show that the topographies for this activity closely match in the pre-response epoch as well. This was to be expected, as the tasks were very similar and there was considerable overlap in the participants whose data was used for analysis. Due to differences in the timing of stimulus onset, direct comparisons have not been made between experiments. Qualitatively, the most notable difference is the more prolonged alpha power decrease in the Contextual Prop experiment towards the end of the trial.

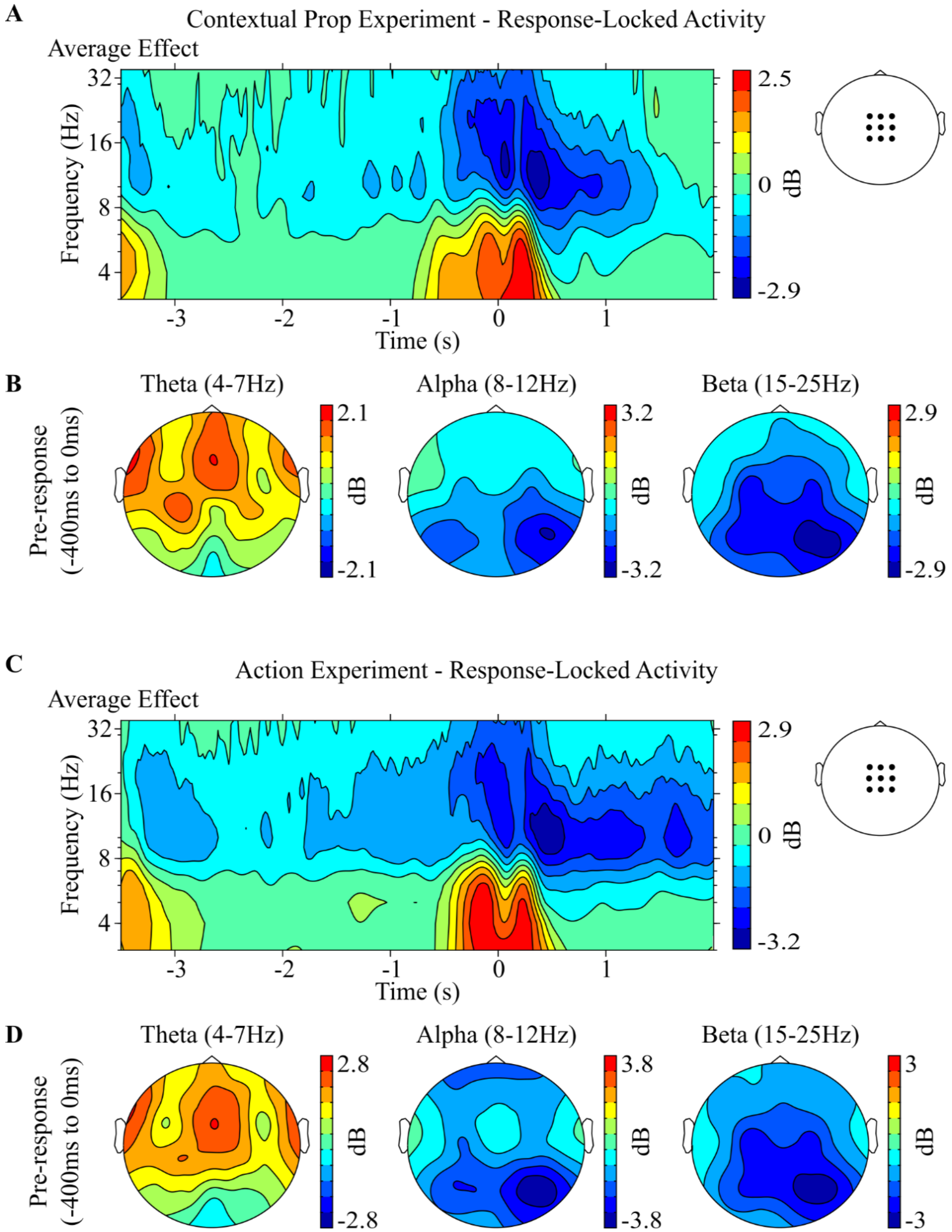


Figure 10. Response-locked time-frequency activity in the Contextual Prop and Action experiments. Data was visualised as a 'sanity check' that the quality was acceptable for further analysis. (A) shows the average effect of ‘shoot’ and ‘don't shoot’ responses in the Contextual Prop experiment. Equal weighting was used for each condition, regardless of the number of trials in each. The average was calculated after baseline correction as the mean of decibel values. Note, the colour axis is formed of

two linear sub-scales: from zero to the maximum value in five steps and from zero to the minimum value in five steps. The sensors represented in the figure are highlighted to the right of the panel. (B) Shows the average effect before response as topographies in theta, alpha and beta. A period of 400ms before response was selected based on average reaction times. (C) and (D) show the same as (A) and (B), respectively, but for the Action experiment.

The baseline period appears to be suitable as very little change in activity (especially in the low frequencies) can be seen in the build up to stimulus onset and participant response. The dominant activity in theta is confined to the frontal midline sensors, centred around position AFz. Note, no data for AFz was actually collected as it was the ground electrode, but the plotting method interpolates between sensors to provide a complete spatial representation of the effect. Simultaneous increases in temporal theta were also observed. Posterior alpha activity can also be seen, with asymmetry biased to the right. As expected, alpha had the greatest variation in power through the course of the experiment. At its lowest point, the mean decrease in alpha power was 3.2dB for the Contextual Prop experiment and 3.8dB for the Action experiment. In beta, the expected decrease in power associated with movement can be seen. However, a larger, unexpected decrease over the posterior sensors, matching the alpha decrease is also present and does not appear artefactual.

4.3.2.3.2. Contextual Prop experiment time-frequency EEG results

Time-frequency data was compared before (-400ms to 0ms) and after (0 to 400ms) responses separately for three frequency bands (theta, 4-7Hz; alpha, 8-10Hz; beta, 15-25Hz). Figure 11 provides an overview of the differences between 'shoot' and 'don't shoot' conditions in the Contextual Prop Experiment. Data was first prepared by averaging the baseline corrected power of trials within the two conditions for each participant. It was then averaged within the frequency band and time period specified by the comparisons, to form six datasets per condition, per participant. Cluster-based analysis was used for all comparisons. A two-tailed paired t-test ($df = 29$) was used as the test statistic for forming clusters. The threshold for belonging to a cluster was set based on an alpha of .05 ($|t| > 2.05$). Cluster significance was determined by the Monte Carlo estimate from 25,000 permutations. For this we also used a two-tailed test with alpha of .05.

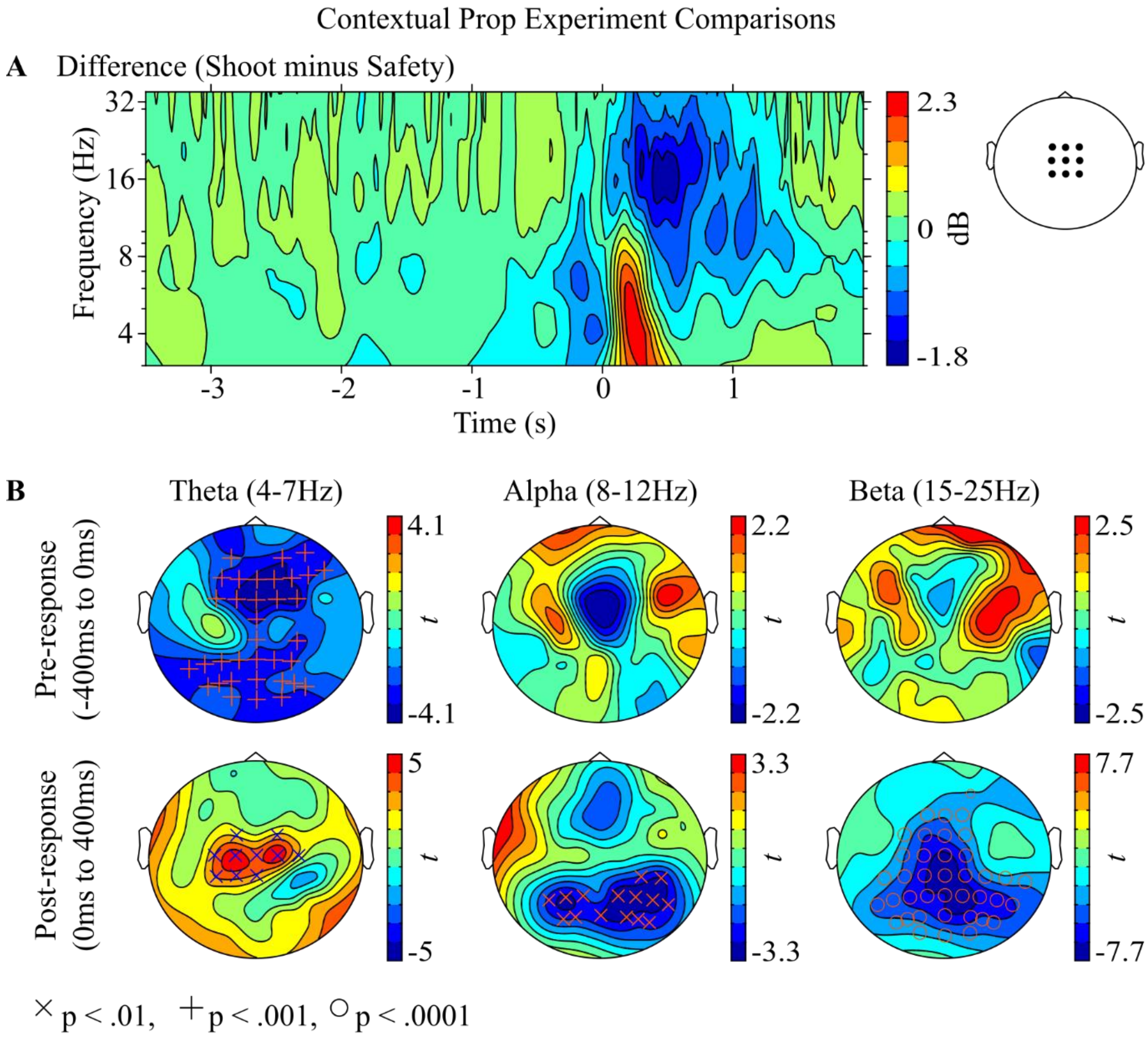


Figure 11. Comparisons of response-locked activity between Contextual Prop experiment conditions. Values show data for the Handgun condition minus the Can condition. (A) Shows the difference in baseline corrected time-frequency data averaged over the central nine electrodes. Note, the colour scale for this plot is made up of two independent, linear scales: one from zero to the maximum dB value and the other from zero to the minimum dB value. (B) The results of cluster-based analysis for data in six datasets. Significant clusters are highlighted in contrasting colours over the sensors within each cluster.

Comparison of pre-response theta activity between the Handgun and Can condition revealed a significant negative cluster distributed over fronto-central and parieto-occipital electrodes, $p < .001$, 95% CI [.0002, .0008]. This suggested significantly greater theta activity in the Can condition in those areas. No other clusters were found in this comparison. No clusters, significant or otherwise, were found between conditions in the pre-response alpha and beta datasets.

We expected greater differences in the post-response time interval as the stimulus varied considerably between conditions at that time. Nonetheless, they were useful to provide context for pre-response activity, as well as forming further ‘sanity checks’. Two positive clusters were found in theta for the post-response interval. One suggested that there was significantly greater central theta power in the Handgun condition than the Can condition, p = .007, 95% CI [.006 .009]. Both conditions had positive theta activity during this time period, relative to baseline. The other positive cluster was not significant, p = .102, 95% CI [.099 .106]. For comparisons in alpha, a significant negative cluster was found over parietal electrodes, p = .01, 95% CI [.009 .011]. The cluster shows that the post-response posterior alpha decrease in the Handgun condition was greater than in the Can condition. In keeping with the average effect, the difference was biased towards the right posterior sensors. A significant negative cluster for post-response beta was also found, p < .001, 95% CI [0 .001]. This cluster was centred over parietal electrodes. Relative to the average effect, it suggested a greater decrease in post-response beta activity for the Handgun condition.

4.3.2.3.3. Action experiment time-frequency EEG results

The analysis in this section is identical to the previous, and most of the observed clusters are similar, so I will be briefer and only key differences will be highlighted. Figure 12 gives an overview of the comparisons between the six datasets of each condition. Again, these are baseline corrected power, averaged within pre- and post-response intervals and the three frequency bands of interest. The only difference in the cluster-based analysis was the degrees of freedom used for the test statistic (df = 26, $|t|$ > 2.06), due to the difference in participant number.

Comparison of pre-response theta activity found a significant positive cluster across left central and temporo-parietal electrodes, p = .028, 95% CI [.026 .03]. Positive clusters were found for pre-response comparisons in alpha, p = .054 95% CI [.052 .057], and beta, p = .085 95% CI [.081 .088], but neither were found to be significant when tested.

The pattern of post-response activity generally mirrored that of the Contextual Prop experiment. A significant positive cluster over central electrodes, extending to temporal areas, was found for comparisons of post-response theta activity, p < .001, 95% CI [<.001 .001]. Post-response beta comparisons revealed a negative cluster across the parietal electrodes, p = .002, 95% CI [.002 .003]. A negative cluster in post-response alpha was found closely matching that topography of the Contextual Prop experiment finding, but was not significant, p = .054, 95% CI [.051 .056].

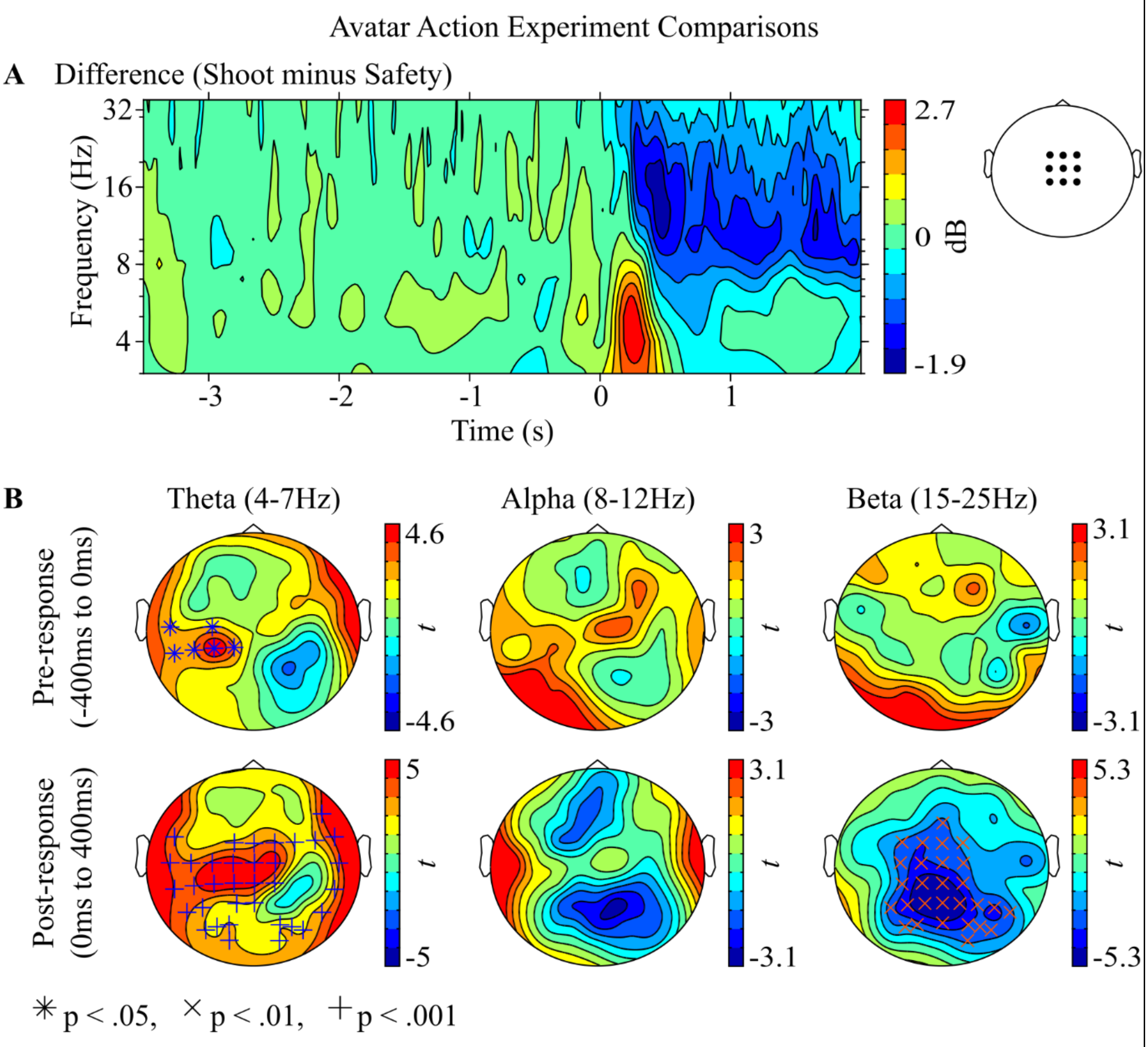


Figure 12. Comparisons of response- locked activity between Action experiment conditions. Values show data for the Attack condition minus the Surrender condition. (A) Shows the difference in baseline corrected time-frequency data averaged over the central nine electrodes. Note, the colour scale is made up of two independent linear scales: one from zero to the maximum dB value and the other from zero to the minimum dB value. (B) The results of cluster-based analysis for data in six datasets. Significant clusters are highlighted in contrasting colours over the sensors within each cluster.

4.3.2.3.4. Both experiments time-frequency EEG results

Our predictions were shared between the two experiments, because their design was very similar. We had not predicted that difference in pre-response theta activity would be related to the source of threat: Action or Contextual Prop. However, findings from both studies suggested that there may be a difference. For this reason, we conducted exploratory analysis to compare the threatening and non-threatening conditions between experiments. Figure 13 shows the results of these comparisons.

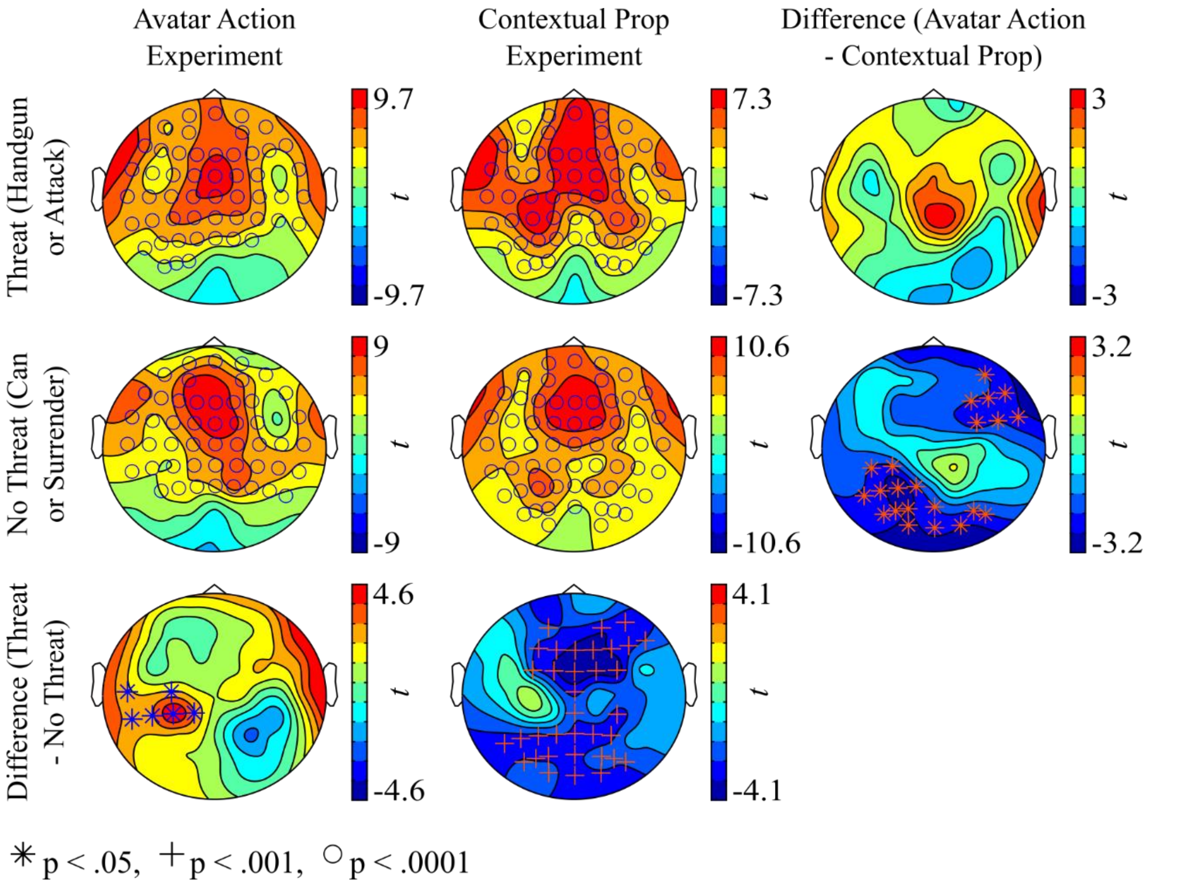


Figure 13. Between-experiment comparisons of pre-response theta activity. The first two columns summarise the main effect we were interested in from both experiments. Where no contrast is stated, differences are shown against baseline. The final column shows the difference between experiments for comparable conditions. Significant clusters are highlighted in contrasting colours over the sensors within each cluster.

Differences against baseline were first calculated. To do this, we used the same cluster-based analysis, using dependent *t*-tests for the test statistic, adjusting degrees of freedom according to the experiment. Relative to baseline, all conditions had a single significant positive midline-frontal and temporal cluster, $p < .001$, 95% CI [<.001 <.001]. They all shared the same statistics, which is why only one is reported. However, as the individual experiment comparisons revealed, their magnitudes and distribution within these clusters varies.

We tested for differences in pre-response theta for threatening and non-threatening conditions between experiments. Despite all participants completing both experiments, exclusion of some datasets did not allow for within-subject comparisons. For this reason, the test statistic used to determine cluster thresholds was an independent *t*-test ($df = 56$, $|t| > 2$). For comparisons across threatening conditions, a single positive cluster was found over central electrodes. This would have suggested the Action experiment had a greater increase in theta activity than the Contextual Prop experiment, but the cluster was not significant, $p = .085$, 95% CI [.081 .088]. For comparisons across non-threatening conditions, two significant negative clusters were found, suggesting greater theta activity in the Contextual Prop experiment. One cluster included occipital and left parietal electrodes, $p = .013$, 95% CI [.011 .014], and the other right frontal and fronto-temporal electrodes, $p = .038$, 95% CI [.036 .041].

4.3.2.6. Source analysis results

4.3.2.6.1. Source analysis motivation

Considering the whole topography (not just the labelled significant clusters) of pre-response theta, the effect in both conditions appears to be extremely similar. The activity highlighted by significant clusters in each is present in the other, just to a lesser extent. This could have been caused by two competing sources or may have been from a single source.

4.3.2.6.2. Frequency analysis for source localisation

Before localising the sources of power in the brain, we needed to analyse the power of signal recorded at the scalp. This required a different procedure than the one described earlier in the time-frequency analysis section because we were less interested in how signal changed over time. Rather, we wanted to measure the coherence between signal recorded at different electrodes for specific frequencies within an epoch.

The same pre-processed data was used as before, but no baseline correction was applied. These data were divided by condition and then into 400ms long epochs pre- and post-response, as well as for the baseline period (3.65s to 3.25s before stimulus onset). PSD was calculated using the Fast Fourier Transform (FFT) for theta (3-7Hz), alpha (8-12Hz) and beta (15-25Hz) for all epochs. Note, theta band was expanded to include activity at 3Hz because the inverse of the central frequency needed to be an integer multiple of the inverse of the sampling frequency. For the frequency range 4-7Hz, the inverse of the central frequency, 5.5Hz, would be 0.1818s, which is not a multiple of $500Hz^{-1}$. To measure coherence, cross-spectral density (CSD) was calculated between all electrodes. Briefly, CSD measures the difference in phase between two signals as the product of their PSD.

4.3.2.6.3. Pre-response theta source analysis results

First, we calculated the source activity for the baseline and activity periods and used cluster-based analysis to find the main sources of activity. As mentioned, this allowed us to compare the direction of effect with the sensor level analysis. It also would aid interpretation of the differences between conditions in terms of their divergence from the main effect. The results of these comparisons can be seen in panel A of Figure 14.

The matched non-threatening conditions were Can and Surrender. In the Can condition of the Contextual Prop experiment, twelve positive clusters were found. One of these was found to be significant, $p < .001$, 95% CI [<.001 <.001]. The positivity and distribution of the cluster approximated observations at sensor level. Notably, the difference against baseline was greatest in the pre-frontal and temporal regions. However, the cluster was large (covering almost the whole cortex), so careful interpretation is needed. For the Surrender condition of the Action experiment, 37 positive clusters were found, one of which was significant, $p = .001$, 95% CI [.001 .001]. Again, the cluster was large, and the greatest differences were in the pre-frontal and temporal regions.

The stimuli for threatening conditions in both experiments were more similar. Both involved being attacked by the virtual human holding a handgun. 146 positive clusters were found for pre-response theta activity with baseline for the Handgun condition of the Contextual Prop experiment. One of these was significant, $p = .007$ 95% CI [.006 .009]. The cluster is more sparsely distributed across the cortex, with no clear source. In the Action experiment, the Attack condition had 89 positive clusters, one of which is significant, p = .001, 95% CI [<.001 .001]. This cluster was distributed, but more clearly centred in the left PFC.

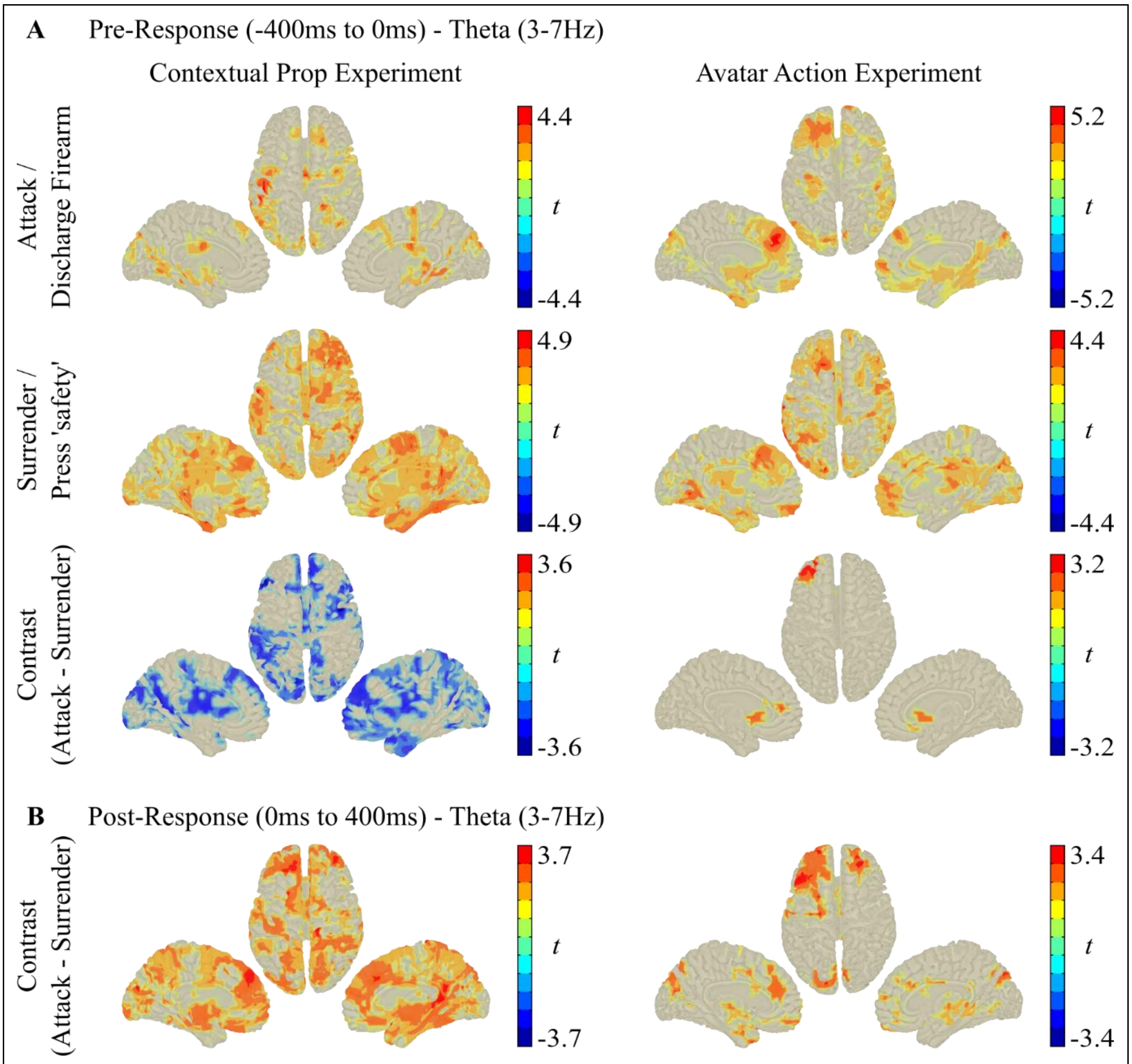


Figure 14. Source analysis findings for comparisons of theta activity in both the Contextual Prop and Action experiments. (A) Shows differences in pre-response theta. The effects of each condition can be seen compared against baseline in the top two rows. The bottom two rows of (A) show the contrast between those baseline corrected effects. (B) Shows differences in post-response theta, but only the result of the contrast between conditions. All values reported are t-values from the cluster analyses described in the main text. Only significant clusters have been plotted.

Overall, the source activity for individual conditions matched the observations at sensor level. This was true for the direction and topography of effect, but also the magnitude. The pattern of greater theta activity in the threatening condition for the Action experiment and the reverse in the Contextual Prop experiment was found. We confirmed this using comparison between conditions at source level. In the Contextual Prop experiment, 83 positive clusters were found comparing pre-response theta activity between Handgun and Can conditions. One of these clusters was found to be significant, p = .003, 95% CI [.002 .004], suggesting greater theta activity in the Can condition. The cluster was distributed, but with the greatest differences in frontal and cingulate cortex. Comparing Attack and Surrender conditions of the Action experiment, a mixture of clusters were found: 98 positive and six negative. One positive cluster was significant, p = .045, 95% CI [.042 .0499]. This cluster appeared far more localised than other clusters reported, but this should not be overinterpreted. It is likely that the cluster just represented a smaller effect, meaning only the greatest differences in the true cluster were highlighted. Nonetheless, it approximately matched the sensor level activity, suggesting the difference in theta came from the left frontal cortex and ACC.

To supplement the pre-response analysis, the same comparisons between conditions were made for post-response theta activity. Comparison of the Handgun and Can conditions of the Contextual Prop experiment revealed 52 positive clusters, one of which was significant, p = .001, 95% CI [<.001 <.001]. It suggested greater post-response activity in the frontal and temporal cortex. For the Action experiment, comparison between Attack and Surrender trials found 131 positive clusters. One significant cluster over the left frontal cortex was found, p = .001, 95% CI [<.001 .001].

4.4. Discussion

4.4.1. Summary of findings

From the behavioural data, we found differences in reaction time depending on whether the virtual human was threatening (Handgun/Attack) or non-threatening (Can/Surrender). As expected, in both experiments, participants were faster to shoot than press Safety. We were unable to evaluate performance by analysing the number of errors in each condition because so few errors were made. Exploratory analysis revealed no significant correlation between age and reaction time and no significant effect of sex in either experiment.

There were many commonalities in event related changes in oscillatory power between experiments. The expected pattern of decreased posterior alpha and increased frontal midline theta power when stimuli were presented was observed in both experiments. When participants were preparing to respond, bilateral decrease in beta power was observed over sensorimotor cortex, typical of induced movement preparation (Wang et al., 2017). Differences between post-response activity between conditions were also highly consistent between experiments: central theta increased, posterior alpha decreased, and central beta decreased. These differences are unsurprising given the differences in post-response stimuli. In the threatening conditions the virtual human was shot and fell, which is considerably more involved than watching a virtual human stand in a surrendering action, as in the non-threatening conditions.

We were most interested in pre-response changes in oscillatory power. These could tell us more about the cognitive function of participants as they observed the stimulus and decided whether to shoot or not. Here, the findings for each experiment diverged. Comparing the Handgun and Can trials of the Contextual Prop experiment, we found greater frontal midline theta in the Can condition. However, running the equivalent comparison in the Action experiment we found greater midline theta in the Attack condition. Despite the differences in significant clusters, the topographies of differences in both experiments were closely matched, which suggested two competing sources with the same approximate effect across experiments.

4.4.2. Interpretation of findings

We believe that having found sensible results that passed our ‘sanity checks’ and generally replicated findings from earlier ‘shoot/don’t-shoot’ experiments, we met our first aim of validating our neuro-VR methods. Specifically, we demonstrated that high-density EEG can be combined with a head-mounted display while participants stand and engage a virtual human. There were some caveats caused by this integration. We found that muscle artefacts were present continuously throughout the data, likely due to the standing position and supporting extra weight from the head-mounted display, EEG cap, and cables. This muscle artefact fell within the typical range (Muthukumaraswamy, 2013) with a broadband elevation from 25-50Hz. Therefore, we could not easily remove it with a low pass filter without removing signals we were interested in. Note that reported time-frequency representations and topographies did not appear to be contaminated by the artefact. This was for two reasons. First, the topography of the muscle artefact was distinct from signal produced from the brain (in particular, over sensorimotor cortex). Second, muscle activity was not event related and so much of it was averaged out in our analyses. Baseline correction then allowed for event related activity to be detected. The main detrimental effect of the muscle artefact on our analysis was removal of other artefacts. Identification of artefacts masked by muscle activity was challenging and the main motivation for the 'pre-cleaning' method described in the previous chapter. An automated method may have removed many more trials than necessary.

The second aim of these experiments was to measure the differences between threatening and non-threatening scenarios. Consistent evidence from previous 'shoot/don't-shoot' studies, supported by our more general understanding of attentional networks, suggested that participants would be quicker to shoot than press Safety, and so we strongly predicted that would be the case in both of our experiments. In the Contextual Prop experiment the average response time was ~510ms and the difference between conditions was ~60ms. This compared with the Action experiment which had an average of ~690ms response time and a difference of ~80ms. These figures closely matched that of other studies (Correll et al., 2002, 2006; Johnson et al., 2018), with the exception of Nieuwenhuys et al. (2012) who reported an average response time of ~850ms with differences of ~600ms between conditions. While neither experiment was a direct replication of any of these studies, we predicted differences more similar to the work of Nieuwenhuys et al. (2012), as we controlled for aim time. With hindsight and reflection on the protocol, it is possible that our participants still took time to aim after a threat was presented, despite the ease of hitting the target.

We did not design the experiments in a way that allowed for direct comparison of response times between experiments. The main reason for this was that the Handgun/Can had to be obscured behind a wall for the first 400ms of the 1s arm raising animation. We considered slowing down the animation such that participants had 1s from the moment the prop was visible until it was pointing at them, but we decided it was more important to control for animation between experiments. These types of compromises are characteristic of naturalistic stimuli, as their continuous properties are more challenging to constrain to discrete events (e.g. animation vs. presenting an image). While we could not directly compare response times between conditions, the differences between them are sensible. In the Contextual Prop experiment, where in the threatening scenarios participants only had 600ms before a Handgun was aimed at them, participants responded quickly (*M* = 480.7ms, *SD* = 62.67ms). In the Action experiment, where they had 1s, they responded slowly (*M* = 651ms, *SD* = 70.24ms). The smaller difference between conditions in the Contextual Prop experiment may have been due to it being more challenging to discriminate between threatening and non-threatening conditions within the more limited time. These differences help to explain some of the variation in EEG recordings we observed between the experiments.

The shifting pattern of theta activity between conditions in the Contextual Prop experiment fit perfectly with our predictions. That is, pre-response frontal-midline theta was greater in the threatening versus non-threatening conditions. Although our methodology favoured analysis of changes in oscillatory power, rather than ERPs, the close link between the N200 and frontal-midline theta (Hajihosseini & Holroyd, 2013) suggests this is a replication of findings from Correll et al. (2006). Our source level analysis supports this interpretation. We found localised differences in pre-response theta to a broadly distributed cluster, but with greatest difference in the frontal and cingulate cortex. Considering the relatively low spatial resolution of EEG (Lenartowicz & Poldrack, 2010), these differences cannot completely support, but are in line with our prediction that activity in the ACC would be necessary to explain differences between the response to shoot or press Safety.

Theta activity in the Action experiment fit less well with our predictions. We did not find a significant difference in frontal-midline theta activity between threatening and non-threatening scenarios. Somewhat in contrast to our predictions, cluster-based analysis revealed greater pre-response theta activity in threatening scenarios. However, over frontal-midline sensors the expected pattern of greater frontal-midline in non-threatening trials was observed but was not significant. Considering behavioural findings alongside this result, it is possible that the manipulation of threat in the Action experiment was not great enough and the task was too easy. If this were the case then our findings would be in line with findings from Go/No-Go tasks. When the Go/No-Go task is made easier, the associated N200 component is later and reduced in amplitude (Benikos et al., 2013; Gajewski & Falkenstein, 2013). Further, in both threatening and non-threatening trials, a handgun was present throughout the whole trial. This may have limited the extent to which Shoot was the prepotent response. Alternatively, predicted effects observed in the Contextual Prop experiment may be specific to object discrimination. We do not believe that this was the case, given the variety of different Go/No-Go style tasks that can elicit N200 and/or difference in theta activity (e.g. Harper et al., 2014; Nieuwenhuis et al., 2003).

4.4.3. Implications for the next experiment

The two experiments presented in this chapter were the precursor for our main experiment investigating the effects of expertise on police decision making. We recognised several improvements and additional considerations related to protocol, experimental design and EEG analysis that could be carried forward.

Regarding protocol, we found that asking participants to aim by bending only their elbow was too constraining and detracted from our aim to reproduce natural behaviour in the lab. Participants still needed to rest their arm between trials to avoid stiffness, so our aim of minimising muscle artefact may have been ineffective, although, as discussed, there was no clear event related muscle artefact. Adding to this, when collecting data from AFOs it would certainly have been a mistake to force them to change the way they aim. We also found that participants were tolerant of the head-mounted display and EEG setup and did not use the breaks often. This meant that our next study could be longer in duration, allowing for a more complex design.

The Action and Contextual Prop experiments had a very simple design. While simplicity was well suited to our early aims, it made the task repetitive and limited analysis. When investigating expertise, it was important that we used a task that was challenging and engaging enough that expertise was required. However, we did not want to have a design where too many comparisons between conditions could be made. This issue will be addressed in the next chapter.

We identified several limitations of our EEG analysis. The length of trials meant that analysis of low frequency oscillatory bands was either limited to the centre of the epoch, as was the case for theta, or unavailable, as for delta. Delta was of particular interest because it has not been studied within a ‘shoot/don’t-shoot’ paradigm, but has been observed during other Go/No-Go tasks (Harper et al., 2014; Kaiser et al., 2019). If we had made the gap between trials longer then this would not have been an issue, but the experiments would have taken longer. For the next study we needed to carefully consider how we could adjust stimulus timings to avoid this issue.

Chapter 5: Comparison of AFO and Novice Performance

5.1. Introduction

The experiments described in the previous chapter demonstrated that we were able to meet the recommendations for research on police put forward by Hope (2016). Specifically, we showed that untrained novices could complete our task and perform in the expected way, meaning that we could control for expertise in the current experiment. We were also able to collect high quality, testable data from realistic, controlled scenarios that were relevant to policing. Having passed these requirements, we were ready to meet the final hurdle of collecting data from police officers. We found this order of development preferable because it meant we could correct any issues identified in our first experiments, before collecting data from police officers. This was important because, as discussed in Chapter 2, there are additional challenges to collecting data from specialist groups like police officers that must be considered (Suss & Boulton, 2019).

We assumed that the selection procedures, training and experience AFOs have been through would mean that they are better than novices at resolving scenarios where the use of firearms is required. For findings from the current study to be meaningful, it would be important to test whether our performance measures were sensitive to this difference (Ejelöv & Luke, 2020). Therefore, if the performance of AFOs in our scenarios was not better than novices it would have been likely that the scenarios and/or performance measures were unsuitable. Any subsequent EEG analysis would then be spurious. This is one of the reasons why having a suitable control group is essential for this kind of research, as, without one, no manipulation check can be made. Our expertise manipulation would be tested by whether AFOs performed better at the task, determined by our performance measures. Given the low number of errors made by novice participants in the previous experiments, we did not expect to find differences in the number of errors between groups. Performance would shift to be measurable by response time instead. Therefore, we predicted that AFOs would have faster response times than novices in all conditions.

Several studies of police performance have benefited from collecting data from police officers, but have not used a control group (Brisinda et al., 2015; Hamilton et al., 2019; Landman et al., 2016a;

Taylor, 2019). Other research has used an alternative method, by collecting data from police officers with a range of expertise and then measuring its effect on their task (Brisinda et al., 2015; Landman et al., 2016b; Mangels et al., 2020; Ward et al., 2011). While effective, such analysis would still benefit from a suitable control group, due to correlations between expertise and other factors, such as age. Further, any research comparing EEG between groups with different aged distributions must consider that age related changes in normal EEG will affect analysis (Knyazeva et al., 2018; Polich, 1997). As we have argued, collecting data from novice participants is generally simpler than collecting data from police officers, so the fact that these studies did not have control groups is likely due to other reasons. For example, infeasibility of novice participants completing the task without firearms training or access to facilities. Here, the benefits of virtual reality can be realised, as there are far fewer limitations on who can use the equipment.

Our prediction that AFOs would have faster response times goes against findings from Johnson et al. (2018). Across two experiments where they tested a total of 151 police officers based in the USA and 220 novice undergraduates, they found that police officers had significantly slower response times in both threatening and non-threatening scenarios. They tentatively suggested that police officers' slower response time was because they were more cautious, as they also made fewer errors. However, they also acknowledged that they did not control for age between groups. The average age of the novice group was 19.2 and, while police officer age was not recorded, they had an average of 8.8 years of experience as police officers. Based on similar demographics reported elsewhere, the police officers' average age would have been approximately 30 (Specialist Arrest Unit; Landman et al., 2016b). Johnson et al. (2018) concluded that any differences in response time may have been due to non-decision related variables, such as increased motor response time, although the relationship between motor response time and age is at a plateau for the age ranges being considered (Pierson & Montoye, 1958). Note, in our earlier Action and Contextual Prop experiments using novice participants, we did not observe a correlation between age and response time within a similar age range. Nonetheless, Johnson et al. (2018) exemplifies the need for suitable control groups when studying expertise so that stronger conclusions can be drawn.

As discussed in Chapter 4, exploratory analysis from our previous studies on novices did not reveal any effect of age or sex on response time. Because of this, for novice participants in the current experiment we did not expect to find any effect of age or sex. However, we were less certain that age would not be an important factor for the AFO group, as their age would likely be highly correlated with their expertise. Given our prediction that expertise would result in faster response times, it would have been reasonable to assume that greater expertise would result in even faster response times. However, we acknowledged that other factors related to expertise, such as caution (Johnson et al., 2018), adaptive flexibility (Boulton & Cole, 2016) and age-related motor response time changes (Pierson & Montoye, 1958) may modulate the relationship between expertise and response time. Therefore, we decided to keep our analysis of demographic variables supplementary to our main analyses. Further, thorough analysis of demographic factors, like age, experience and sex, was outside the scope of our research. To do so would require a larger study with many more expert participants, including those before and during training.

One aim of the current experiment was to make the task more complex and naturalistic, versus our earlier studies. When deciding how to make these changes, we relied again on feedback from firearms instructors. As before, threat level and compliance remained the most important variables, so rather than testing them independently, we decided to combine them. In doing so, we would make the task more challenging and therefore more sensitive to expertise. Another critical piece of feedback related to the use of less-lethal force. Following the United Nations Basic Principles on the Use of Force and Firearms by Law Enforcement Officials (Office of the High Commissioner, 1990) applied in UK law, AFOs should be equipped with a range of weapons to allow for differentiated use of force (College of Policing, 2020). For this reason, AFOs in the UK are typically deployed with, at minimum, a sidearm, such as a Glock pistol (Glock Ges.m.b.H, Deutsch-Wagram, Austria), and a less-lethal firearm, such as an X26 Taser (Axon Enterprise, Inc., Scottsdale, AZ, USA). These firearms allow them to use reasonable force across a range of different situations. To improve realism, we decided to incorporate the use of multiple firearms into our scenarios. The details of how we implemented this will be described in the methods section. In brief, a participant’s decision to use either a Glock, Taser or no firearm would be based on the level of threat. Their later decision to shoot or not would be based on the virtual humans’ compliance.

The addition of a second decision regarding firearm selection and use of reasonable force would impact other areas of the experiment. First, we had to consider whether novice participants would be able to complete these more complex scenarios. We reasoned that in a virtual reality simulation the use of a Taser is identical to the use of a Glock, therefore novice participants would still be able to complete the task effectively. Second, the introduction of sequences of decisions effected both experimental design and subsequent analysis, because the firearm decision would affect the later 'shoot/don't-shoot' decision. This meant that rather than a standard factorial design, independent variables would be manipulated in sequence. There are several benefits to this approach. It would allow us to disambiguate behaviours and neural correlates related to processing of threat level and compliance. Further, sequences of actions, where earlier decisions impact later ones, are a feature of naturalistic behaviours (Sonkusare et al., 2019), and so would help us work towards our aim of collecting ecologically valid data.

Despite these changes, many aspects from the Action and Contextual Prop experiments can be found in the current experiment. Therefore, we expected to find similar patterns of within-subject behavioural differences. We were confident that both experts and novices would be faster to shoot than press Safety, regardless of the firearm type equipped. We expected that faster response to threat would extend to the earlier preparation stage as well. Specifically, response times should decrease as the threat level increases, so participants would be quickest when equipping a Glock, followed by the Taser, followed by no firearm. One additional test, not available in the earlier experiments, was whether there would be any difference in response time between discharging a Glock or a Taser. At that stage in the scenario it was unclear which condition was the more threatening of the two, if either. Participants would be suitably equipped to apply reasonable force in both. Therefore, we did not predict any difference in reaction time between these two conditions. More likely, sensitivity to threat, measured by difference in response time, would be found when participants initially equipped the firearm, because, at the distance the virtual human would be from the participant, a handgun would be more threatening than a knife. We were particularly interested in differences between response times to equip a Glock versus a Taser. As both events were equally likely, any difference in reaction time would be due to a combination of bias in preparation (i.e. which holster was the participant's hand closest to at the start of the trial) and orientation towards threat.

We also expected similar patterns of within-subject differences in EEG across all groups. Any differences in between-group analysis would be differences of degree and not of kind. The ‘shoot/don’t-shoot’ phase of the current experiment would be closely matched with the main task of the Action experiment as the participant’s decision would be based on the animation of the virtual human. Due to these similarities, we predicted that within-subject differences at the ‘shoot/don’t-shoot’ decision stage of the current experiment would replicate the Action experiment. As discussed in Chapter 4, our findings from the Action experiment did not fit perfectly with our predictions. There was some evidence that when participants had to inhibit their prepotent response in the non-threatening condition, greater frontal-midline theta was observed. However, alongside this smaller effect, we observed significantly greater theta activity across left temporal and central electrodes in the threatening scenarios. It was unclear what the source of this difference was. If replicated, the graded threat level would help us investigate this further.

The first decision regarding whether to equip a firearm and, if so, which one, was not a direct replication from the earlier studies. However, the stimuli used at this stage did closely match the Contextual Prop experiment, so we expected similar results. We predicted that when participants decided not to equip a firearm, there would be a greater increase in frontal-midline theta activity, suggesting response inhibition. As mentioned regarding the behavioural data, the additional actions participants could take in the current study allowed us to investigate response to threat more thoroughly. We were especially interested in comparing pre-response EEG for when a Glock versus a Taser was equipped. The stimuli in these conditions would be well controlled, so any differences would be due to perception of threat. Here, we expected to find a difference between novice and AFO participants. We hypothesised that AFOs would recognise the importance of different levels of threat on their actions as this is part of their training. To our knowledge, greater sensitivity to threat related to expertise has not been investigated before and so that analysis would remain exploratory.

Some investigations of the effects of expertise on police decision making have been conducted. Johnson et al. (2014) aimed to identify electrophysiological indicators of expertise in police firearms decision making. They presented participants with a series of video simulations which could be treated as a ‘shoot/don’t-shoot’ task. Participants were equipped with a modified firearm that provided tactile feedback and allowed for measurement of where the shot would have hit in the scenario. If the shot hit the actor, then the software presenting the scenario would cut to another video where the actor was shot. A great amount of work is required to produce scenarios of this kind and, while they are effective, they demonstrate the benefits of virtual reality, where software can update stimuli on-line with little effort. The researchers tested 24 participants: six police officers; six military personnel; and twelve novices. During each scenario they measured performance as 'pass' or 'fail' based on the 'shoot/'don't-shoot' decision, response time, and accuracy of the shot. Note, they did not directly report response times, but they did find that experts made significantly fewer errors, according to their performance metric. Throughout each scenario they also recorded 8-channel EEG. The researchers extracted several metrics from 'pass' trials in which participants discharged their firearm. Of interest to our research, they measured differences in frontal-midline theta and change in alpha between the novice and expert (police/military) groups. They found that experts had significantly greater frontal-midline theta and a greater decrease in alpha power versus the novice group. While identifying these measures may have been beneficial to their aim of objectively measuring expertise, for our purposes only limited conclusions could be drawn regarding the source of these differences. This was because no comparisons were made between conditions. Further, there are several limitations to the EEG analysis methods used. It is unclear whether relative or absolute theta power was compared between groups, but the reported values (μV^2) suggest absolute power was used. Multiple factors, including differences in impedance, could make differences in absolute power arbitrary (Ferree et al., 2001). Power was also calculated for the whole scenario, so it is unclear whether differences between groups are event related or due to individual differences. Even so, this study, to our knowledge, is the first and only study that has investigated differences between police officers and novices using neuroimaging. In the current study we aimed to build on their work using our neuro-VR methods that would allow us to analyse event-related changes in oscillatory power.

The effects of expertise have been investigated using paradigms unrelated to police decision making. Athletes in open skill sports (i.e. sports requiring dynamic reactions; Poulton, 1957), such as fencing, taekwondo, table tennis, basketball and baseball, have been shown to have faster response times in Go/No-Go tasks versus non-athletes (Brevers et al., 2018; Bruna et al., 1992; Di Russo et al., 2006; You et al., 2018) and athletes in closed skill sports, such as swimming (Wang et al., 2013). Matching our assumptions about AFOs, these studies have assumed that the selection procedures, training and experience of open skill athletes has resulted in better response inhibition and stimulus discrimination. This conclusion is supported by analysis of neural correlates. Di Russo et al. (2006) collected EEG for athletes and non-athletes for a Go/No-Go task and a simple reaction time task. They then calculated the difference in ERPs from Go and No-Go conditions, to isolate correlates of response inhibition. In addition, they calculated the difference between ERPs from the Go condition of the Go/No-Go task and ERPs from the simple reaction time task, as a measure of stimulus discrimination. Using a dipole fitting approach, they compared differences in visual perception and response inhibition between groups. Analysis of differences in visual perception revealed earlier activity in the posterior cingulate cortex in the athlete group. Activity related to response inhibition was localised to the PFC and ACC, where athletes showed greater increases in activity. By relating the time course of the differences in ERP waveforms, they related activity in the posterior cingulate cortex to the N100 component and activity in ACC and PFC to the N200. This second finding has since been replicated, with the additional outcome that the N200 peak is also earlier for athletes (You et al., 2018). If these findings for open-skill athletes translate to AFO expertise, then they can be used to make predictions about between-subject differences in the current study. We should observe greater increases in frontal-midline theta activity for AFOs versus novices when they inhibit responding to threat.

In summary, in the current study we aimed to develop our neuro-VR ‘shoot/don’t-shoot’ paradigm into a more realistic task. Participants would be presented with a series of scenarios where they would have to identify and respond to varying levels of threat and, if necessary, apply reasonable force. Crucially, we would collect data from expert AFOs to compare against novices. In our analysis, we would first check that our expertise manipulation was successful, based on differences in behavioural data. Following this, we would test a series of hypotheses related to within- and between-subject differences in EEG. Our predictions about within-subject differences were based on our earlier findings from the Action and Contextual Prop experiments. In addition to these, some new tests related to identification of varying threat levels would be introduced. The most critical part of our analysis would test the effects of between-subject differences in expertise. Only limited research has been conducted in this area, but we relied on research on open skill athletes as analogous to police officers to make our predictions

5.2. Method

5.2.1. Research design

There were three independent variables in this experiment. Expertise varied between-subject and was determined by the group participants belonged to: AFOs, Matched Novices or Younger Novices. Two within-subject variables determined the stimuli used in each trial. The first of these was the Contextual Prop held in the virtual humans' hand. It could be either a Can, Knife or Handgun. The second, Action, determined whether the virtual human decided to Attack or Surrender.

5.2.2. Participants

We collected data from the three groups of participants concurrently with two aims in mind: first, we wanted to control for age and sex between the AFO and Matched Novice groups; second, we aimed to match the sample size of the novice groups to that of the AFO group. Figure 15 gives an overview of participant descriptive statistics. The same exclusion criteria were applied as for the earlier experiments to ensure participant comfort and good data quality. This research was given a favourable opinion by a university research ethics committee (ref 1179).

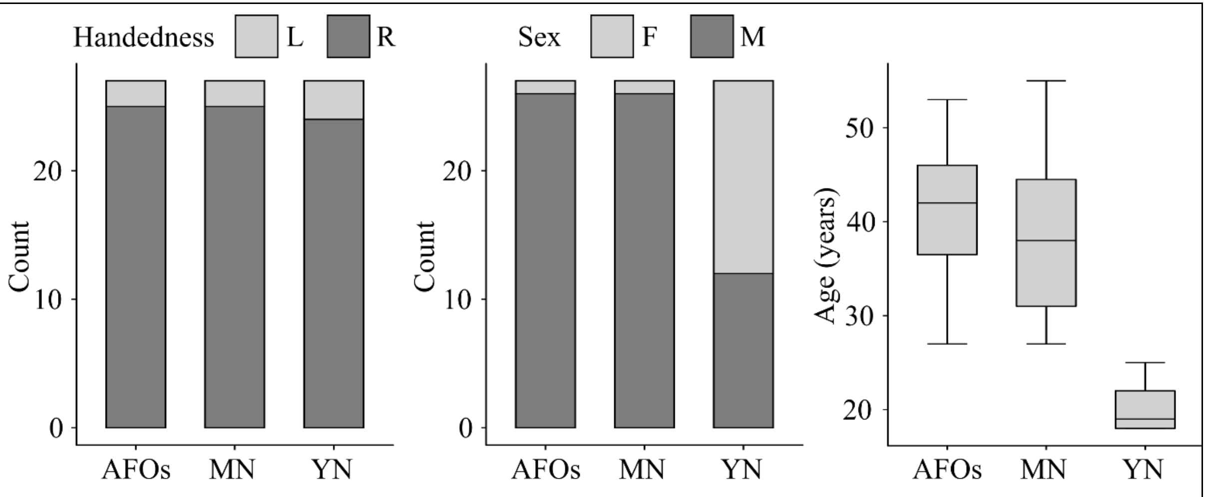


Figure 15. Participant descriptive statistics for the AFO, Matched Novice (MN) and Younger Novice (YN) groups. These statistics show that we succeeded in sampling a matched control group for comparison with the AFO group. Comparisons between the AFOs and Matched Novice control group for the measures we collected show no significant differences between the groups. The Younger Novices group had no overlap in age with the other groups.

The most recently available statistics on UK Police officer diversity state that women account for 30.4% of police officers (Home Office, 2019b). However, this value drops to ~5% in firearms and tactical roles, which includes AFOs (Home Office, 2010) –although this figure is less recent. Therefore, we expected most of our AFO participants to be men. Further, we were advised by our collaborators in the police that to become an AFO several years of experience are required and that few, if any, participants would be under 25. We used this guidance when recruiting participants for the novice groups and adapted as we collected more data from AFOs.

During three week-long visits to the Tactical Training Centre we collected data from 27 AFOs employed by either Durham Constabulary or Cleveland Police. All of these officers were fully trained as AFOs. Some were trainers, commanders or worked in specialist roles, but all could be deployed in the capacity as an AFO and were up to date on training. Their age ranged from 27 to 53 (M = 40.6, SD = 6.8) and 26 were men. Their experience as police officers ranged from five to 32 years (M = 17.1, SD = 6.9) and they had been trained as AFOs for between one and 22 years (M = 10.6, SD = 7). All officers were on duty and volunteered their time and gave their informed consent to take part in the experiment.

Participants in the novice groups were recruited from various sources and included members of the public, students and staff at Aston University, and staff at the University of Nottingham. Along with the other exclusion criteria, these participants must have had no experience working for the police or armed forces. They were offered either £20 or, if applicable, course credits for their time. All gave their informed consent to participate in the experiment.

For the Matched Novices group, we collected data from 29 participants. Two datasets were removed before data collection was completed (explained in the behavioural data preparation section), leaving 27 participants in this group. Their age ranged from 27 to 55 years (M = 38.3, SD = 8.9) and 26 of them were men. We did not attempt a one-for-one match with the AFO group for age, but instead aimed for similar distributions. The difference in age between the female participants in the Matched Novice and AFO groups was three years. By chance, both groups had two left-handed male participants.

Our intention was to collect a third set of data more typical of a psychology or neuroscience experiment. However, our efforts to form the Matched Novice group meant that this group was younger than it would have normally been, hence the name Younger Novices. We collected data from 30 participants and excluded three from analysis during data collection. The 27 remaining participants' ages ranged from 18 to 25 (M = 20.4, SD = 2.6). There were twelve male and three left-handed participants.

Like our earlier experiments, we did not pre-determine our sample size using a power calculation. As with many studies conducted based on sparse previous research, there were other overriding factors that informed our sample size (Bacchetti et al., 2011). Specifically, we were collecting data from a limited population police officers at a remote facility. Therefore, our aim was to collect as many expert AFO participants as possible during our visits and then match our control group sample sizes to them. If a power calculation were possible and suggested a larger number of participants then we would likely have rejected it after consideration of more practical issues and so it was redundant.

Overall, we achieved our aims for data collection and were able to analyse three groups of 27 participants. To do this, data was collected at three sites, highlighting a benefit of this kind of research: you can take your virtual lab with you. The same equipment was used, and the same researcher collected data in all instances.

5.2.3. Materials and apparatus

5.2.3.1. Virtual reality and EEG setup

The same virtual reality and EEG equipment and setup was used as for the earlier experiments. Briefly, an Oculus Rift CV1 head-mounted display presented the experiment and participants responded using two Oculus Touch controllers held in their hands. A waveguard 65-electrode (Ag/AgC) EEG cap (ANT Neuro, Hengelo, Netherlands) was worn underneath the head-mounted display. A structured light sensor was used to create 3D scans of participants' heads while they were wearing the EEG cap, following our established protocol.

5.2.3.2. Virtual environments and virtual human

The virtual environment and virtual human created for this experiment closely matched our earlier experiments. They were produced using our established development pipeline. The environment was made up of two walled courtyards separated by another wall with an entrance in the middle. Throughout the experiment, participants would stand in the centre of one courtyard, facing the entrance. From the participant's perspective, the virtual human started each trial in the opposite courtyard, always on the right side of the entrance. For the same reasons as the previous experiments, a single virtual human was used for all trials. We updated the virtual human, but he looked quite like the previous one: a white male casually dressed and wearing a neutral expression. The reason for the update was to improve compatibility with a new animation pipeline.

5.2.3.3. Action mapping

Participants needed to use their virtual hands to complete the task. Four buttons on the hand controller (they were instructed not to use any of the other ones) allowed them to do this: a trigger for the index finger, a trigger for the middle finger, and two buttons for the thumb. The thumb buttons were very close to each other, so were mapped to the same function to avoid confusion. The middle finger trigger was used for grabbing objects, the index trigger for discharging firearms and the thumb buttons for pressing Safety and indicating readiness to continue. The two triggers could only be used on the dominant hand controller. The design of the controllers meant that these actions were quite natural.

Two virtual holsters were bound to the participants' movements which allowed them to grab either a Glock or a Taser. The Glock was the sidearm and placed close to the hip on the dominant side. The Taser was placed at the centre of the chest. We were advised by firearms instructors that these are common placements for AFOs, although they may vary with preference and task. Placement was only approximate as we had no tracking information for the chest or hips. Instead we used the head position and assumed the chest was just below the head and the hips further below and to the sides. For the most part, this worked well as participants' posture changed very little. It also ensured that participants did not look down when equipping a firearm,. This would have been a problem, as the holster was bound to their head position and assumed to be relatively stationary. We wanted to avoid this because it meant they might miss some stimuli and/or introduce muscle artefacts into the EEG data. One issue was that we did not adjust for height, meaning the ‘hip’ was slightly too high/low for some participants. This was not a significant problem but should be considered in future research.

5.3.4. Procedure

5.3.4.1. General procedure

We attempted to treat all participants in the same way, regardless of expertise, and all were given the same written instructions. Participants were told that the session would take approximately two hours, allowing time for the practice, EEG setup, experiment and hair washing. The experiment itself would take 35 minutes. After reading the information sheet and asking any questions they had, the head-mounted display was setup and adjusted for comfort and clarity of display. They then practiced the task under instruction until familiar with the buttons and instructions. EEG was then applied, and the participant was coached through minimising artefacts by demonstrating the effect of talking and moving muscles on the recording. They then completed the main experiment which was made up of ten blocks of 20 trials. Between each block participants were offered a short break in which they could sit down and take the head-mounted display off. This break was rarely accepted and most participants favoured completing the experiment without a break.

5.3.4.2. Task procedure

Each trial had two stages, Preparation and Decision. The Contextual Prop condition determined the stimulus at the Preparation stage and the Action determined the stimulus at the Decision stage. Participants made one response per stage in most instances. Figure 16 presents the timings and structure of possible trials in detail. For reference, Figure 17 shows an example of the stimuli used in one scenario.

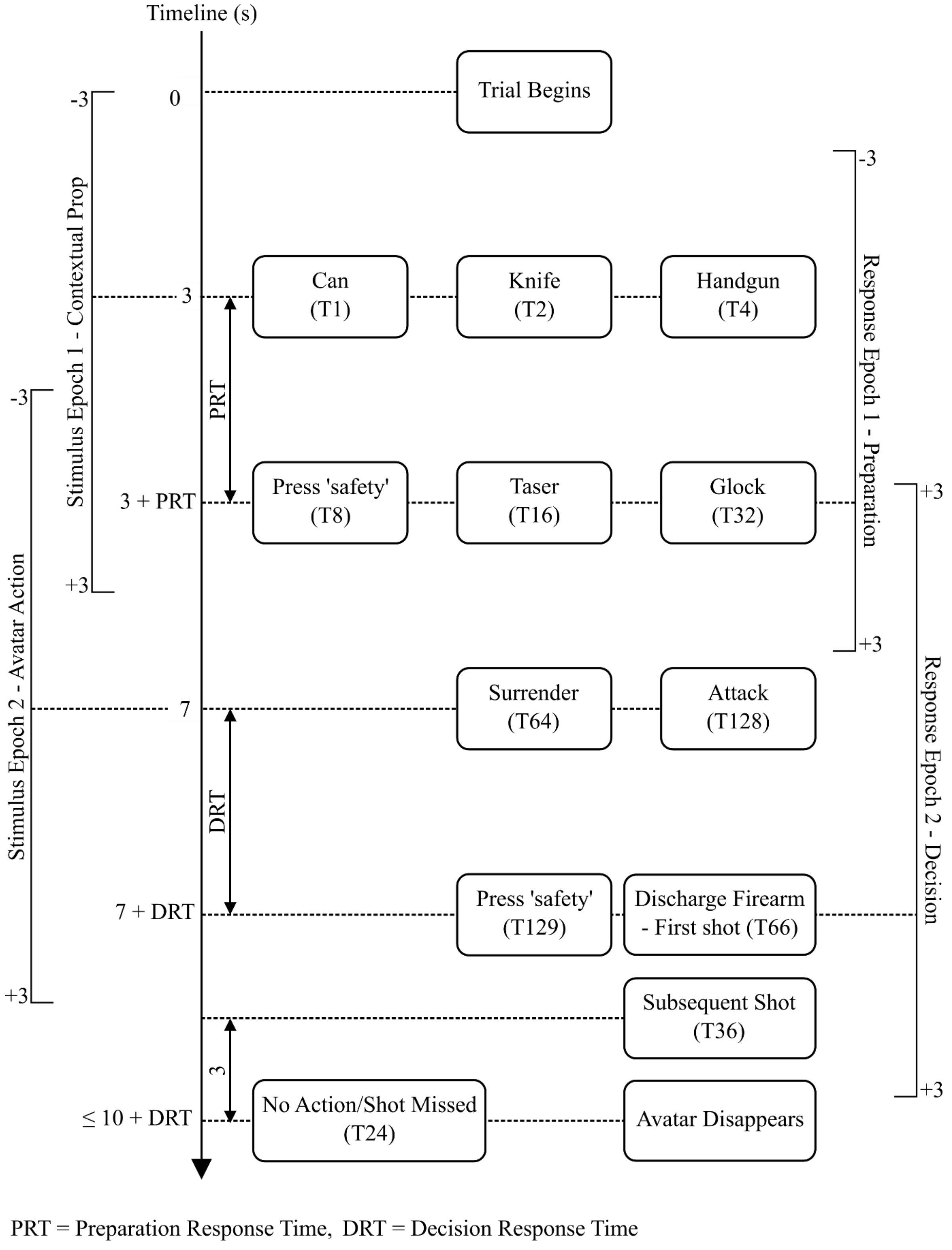


Figure 16. Overview of experimental design and timings. After a trial began, a Contextual Prop held in the virtual humans' hand was presented. If the virtual human held a Can, he would always Surrender. If the virtual human had a Knife or a Handgun, he would either Surrender or Attack the participant. The timings of these events allowed for a minimum of three seconds either side for epochs in the EEG data to be taken. In addition, if the participant missed their first shot, or any subsequent

shots, the end of the trial was pushed back so that their final shot still had a three second break before the end of the trial. In brackets below each event is the code used, where necessary, to identify the event in the EEG recordings.

At the Preparation stage, the virtual human would walk from his starting position behind the wall (completely obscured) to stand at the entrance in front of the participant. The Contextual Prop condition determined the item held in the virtual humans' hand and would first appear as he came from behind the wall, into view. Participants would prepare themselves to deal with the threat, if present. They had three possible actions to choose from at the Preparation stage: equip their Glock, Taser, or press Safety. Their instructions told them that the correct preparation for each Contextual Prop was to grab the Glock for Handgun, Taser for Knife and press Safety for Can. Whichever action was completed first was the final action for that task. When Safety was pressed a 'click' sound was made and the firearms holsters were disabled, preventing either being equipped. When a firearm was grabbed it was bound to the virtual hand until the end of the trial. It could not be dropped or replaced with the other firearm. Participants were instructed to aim the firearm at the centre of mass of the virtual human, in preparation for the next stage.

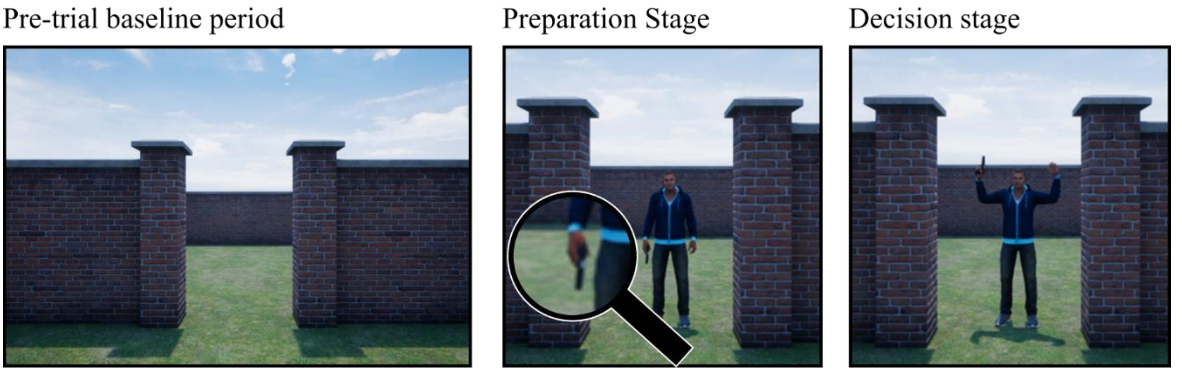


Figure 17. Examples of stimuli from a scenario where the Contextual Prop is a Handgun and the Action is to Surrender. Note, these stimuli were presented on a head-mounted display, so the precise presentation of these stimuli varied continuously with participant movement.

The Decision stage began when the virtual human either Attacked or Surrendered, as determined by the Action condition. The Surrender animation was the same regardless of Contextual Prop; the virtual human would raise both hands while standing in place. He did not drop the prop before doing so, although we would change that if running the experiment again. The Attack animation for the Handgun raised the virtual humans' right arm to point it at the participant. For the Knife, the same arm animation was used, raising it to the participant, but the virtual human also ran towards the participant. No Attack animation was needed for the Can condition because the virtual human never attacked with it. All animations took one second to complete. The participants could decide to either shoot the virtual human or press the Safety button. Whichever action they chose disabled the other. They were instructed to press Safety as soon as they saw the virtual human surrendering. A click sound gave them feedback to let them know the button had been pressed. Likewise, they were to shoot as soon as the virtual human attacked. If they missed their shot, they were able to take another one. When the virtual human was shot his animation blended into a 'shot' animation and he fell to the ground, just as for the previous experiments. When the Safety was pressed the animation continued to its end. At the end of a trial the virtual human disappeared, and any equipped firearms were automatically replaced in their holster.

5.3. Results

5.3.1. Behavioural results

5.3.1.1 Data preparation

The established procedure of importing log files produced by the game engine after each trial into MatLab for data preparation was used. Relative to the earlier Contextual Prop and Action experiments, considerably more behavioural information was collected each trial. In addition, the more complicated, two-stage design meant that evaluation of performance for analysis and outlier detection purposes required more careful coding of correct, incorrect and 'no action' trials.

Nonetheless, evaluating performance at the Preparation stage was simple: participant response was considered correct if the firearm equipped, Taser or Glock, matched the level of threat, Knife or Handgun, respectively. If the Safety button was pressed when the virtual human was non-threatening, indicated by them holding a Can, this was also considered a correct response. Note, this is distinct from when nothing was equipped, and Safety not pressed. If that were the case, then the trial was coded as 'no action'. Incorrect preparations included equipping the wrong firearm for the level of threat presented in the scenario or equipping either firearm in the Can condition.

Coding for performance in the second, Decision stage, where participants decided to Shoot or press Safety, depending on the Action condition, was more challenging. A correct response was recorded if it was proportionate to the virtual humans' action. i.e. in the Attack or Surrender condition, the correct response would be to Shoot or press Safety, respectively. An incorrect response would be the alternative action, and no response at all was considered a 'no action' trial, just as for the Preparation stage. However, response at the Decision stage was dependent on the earlier response at the Preparation stage and so could not be considered in isolation. This was because if at the Preparation stage the wrong response was made, then a proportionate response at the Decision stage was not possible. For example, if a participant mistakenly equipped their Glock in the Knife condition then they could not make a correct response, following task instructions. Participants knew that this was the case and so measurements of behaviour and EEG may have been affected. Therefore, if this happened in a trial, regardless of the participant's response, the trial was removed from analysis and did not contribute to the total number of trials from which correct and incorrect response percentages were calculated.

To determine whether any participant datasets should be rejected as a whole, the same principles were applied as for the previous experiments. To recap, many repetitions of each condition were required for the EEG data to be amenable to analysis. We also wanted to remove any participants who had misunderstood the task. Inspection of the number of incorrect responses revealed that no participant made a large number of errors; the highest percent incorrect for an individual was 6%, and the highest percent incorrect for a single condition (by the same individual) was 25%. However, for some participants there were many trials in which no response was given. Most commonly, this happened when the Safety button should have been pressed, which was unfortunately not detected during the experimental session. This problem was realised partway through data collection and participants who had a large proportion of 'no action' trials were removed so that their data could be replaced with data from additional participants to achieve our aims for data collection. As this process needed to take place during data collection, the threshold for removal could not be based on the distribution of the total sample. However, the percent of 'no action' trials for the removed participants was greater than three standard deviations above the mean of the final sample. In total, five participants' datasets were removed and replaced: two from the Matched Novices group and three from the Younger Novices. No participants were removed from the expert AFO group.

Unlike the earlier experiments, trials with outlier response times were not removed from analysis. This was because we allowed greater pauses between trials and responses within trials, removing the issue of overlapping response-locked epochs. To calculate average response time for each condition the median average of correct trials was used to mitigate the effect of outliers without removing data. In all conditions and across all three groups, participants' median response times were within three standard deviations of the mean and so none needed to be considered for removal.

To assess performance across conditions and between groups, we used the ratio of incorrect trials to the total number of trials in which a response was made. This allowed performance in the Decision stage to be calculated without undue influence from performance at the Preparation stage. However, on average, this ratio accounted for only two percent of trials. We felt that this meant no meaningful conclusions about performance could be made using this metric. It is likely that, due to the relatively unrestricted time allowed for response, there was little trade-off in accuracy for speed (Wood & Jennings, 1976). Therefore, only response time was used to assess performance.

5.3.1.2. Behavioural data results

5.3.1.2.1. Data preparation

Response times from the Preparation and Decision stages were analysed separately as the measurements were independent of each other and belonged to different groupings of conditions. This also simplified the design, because the Preparation stage Can condition did not require a response at the Decision stage, and so there would have been an empty field in the analysis. Therefore, the data were divided into two datasets and were exported from MatLab into R: A Language and Environment for Statistical Computing (R Development Core Team, 2019) for analysis. For a graphical overview of the results discussed in the next two sections, see Figure 18.

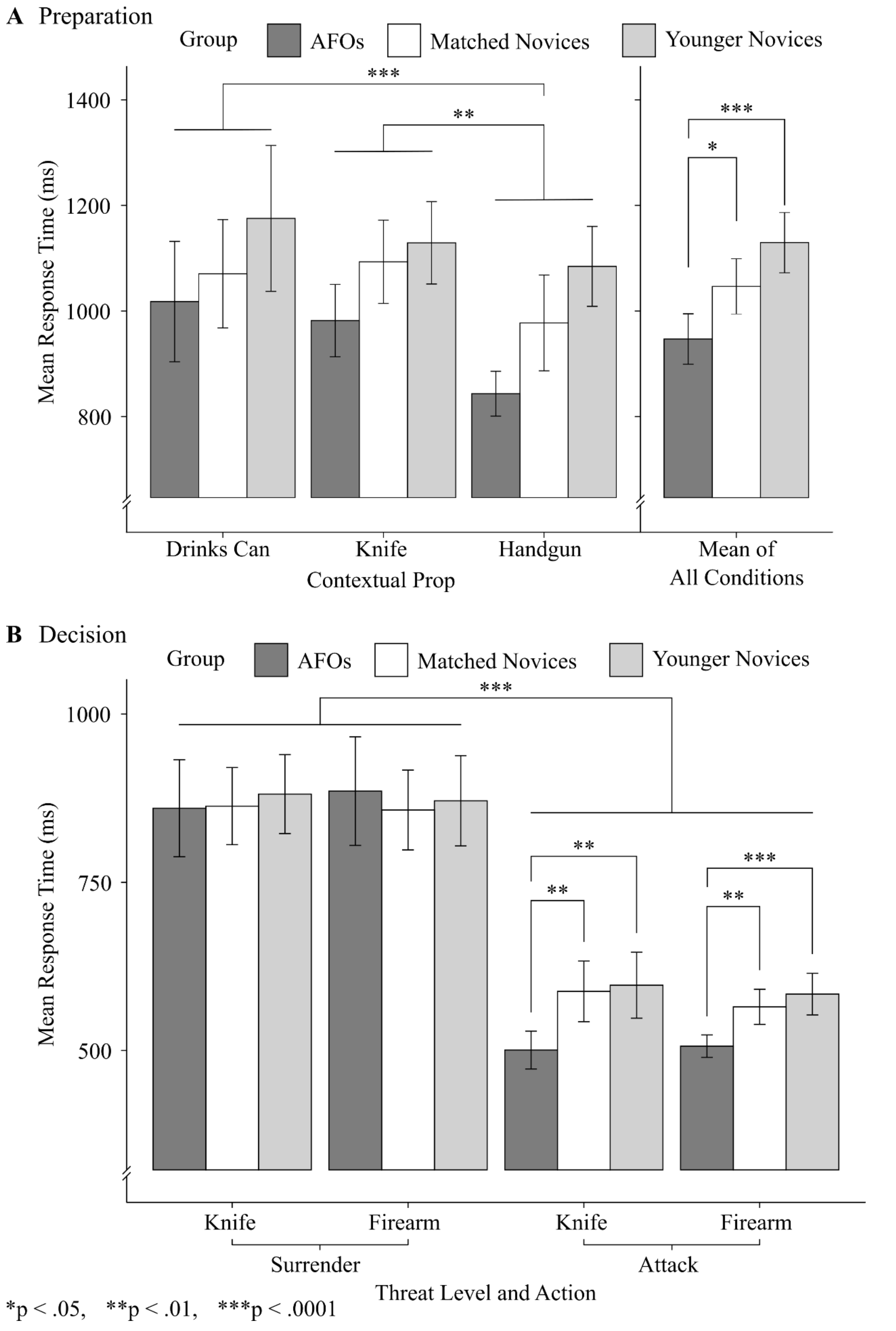


Figure 18. Summary of differences in response times between conditions and groups. The key finding was that AFOs were faster to shoot than both novice groups. Connecting lines mark significant differences between the bars they cover. (A) Shows the results from the Preparation stage analysis of

variance. Within-subject comparisons can be seen to the left with additional pairwise comparisons for the Contextual Prop independent variable. To the right, between-subject comparisons are shown. (B) Results from the Decision stage analysis of variance are shown alongside pairwise comparisons within the Attack condition of the Action independent variable.

5.3.1.2.2. Preparation stage

A three (between-subject, Group: AFOs vs. Matched Novices vs. Younger Novices) by three (within-subject, Contextual Prop: Can vs. Knife vs. Handgun) mixed factor analysis of variance was used to test our hypothesis that performance, measured by response time, would differ between groups and conditions. A significant main effect of Contextual Prop on response times at the Preparation stage was observed, $F(2,156) = 8.36$, $p < .001$, $\eta_{ges}^2 = .05$. A significant main effect of Group was also found, $F(2,78) = 8.28$, $p < .001$, $\eta_{ges}^2 = .10$. There was no significant interaction between Group and Contextual Prop, $F(4,156) = 0.69$, $p = .6$, $\eta_{ges}^2 = .01$.

As no significant interaction was found, the direction of the main effects was tested. The results of pairwise comparisons of the Contextual Prop conditions can be seen in Table 3. These suggest that the main effect of Contextual Prop was driven by faster response times in the Handgun condition only.

Table 3. Pairwise comparisons of the Contextual Prop conditions. Dependent *t*-tests were calculated between each condition and the *p*-values corrected using the Bonferroni method. Cohen's *d* is also reported as a measure of effect size.

Contextual Prop Condition	*M* (ms)	*SD* (ms)	Bonferroni correct *p*-values / Cohen's *d*		
			Can	Knife	Gun
Can	1088	304.8	-		
Knife	1068	198.3	1 / 0.08	-	
Handgun	968	206.4	.004* / 0.47	< .001* / 0.49	-

Note, * marks significance.

Similar pairwise comparisons for Group can be seen in Table 4. These results show that the observed main effect of Group describes AFOs as significantly faster than the Younger Novices and

Matched Novices. No significant difference was found between Matched Novices and Younger Novices, although the effect size was moderate.

Table 4. Pairwise comparisons of the Contextual Prop conditions. Independent *t*-tests were calculated between each condition and the *p*-values corrected using the Bonferroni method. Cohen's *d* is also reported as a measure of effect size.

Group	*M* (ms)	*SD* (ms)	Bonferroni correct *p*-values / Cohen's *d*		
			AFOs	Matched Novices	Younger Novices
AFOs	948	139.8	-		
Matched Novices	1047	181.6	.023* / .61	-	
Younger Novices	1130	169.3	< .001* / 1.17	.079 / 0.47	-

Note, * marks significance.

5.3.1.2.3. Decision stage

A three (between-subject, Group: AFOs vs. Matched Novices vs. Younger Novices) by two (within-subject, Contextual Prop: Knife vs. Handgun) by two (within-subject, Action: Surrender vs. Attack) mixed factor analysis of variance was conducted. As expected, a main effect of Action was found, $F(1,78) = 315.81$, $p < .001$, $\eta_{ges}^2 = .59$. The effect size was large and matched the findings from our earlier experiments, suggesting faster response times when under threat. No significant main effect was found for Contextual Prop, $F(1,78) = 0.31$, $p = .58$, $\eta_{ges}^2 < .01$. The following interactions were not found to be significant and had very small effect sizes ($\eta_{ges}^2 < 0.01$): Group by Contextual Prop; Contextual Prop by Action; Group by Contextual Prop by Action.

We had predicted that AFOs would be faster to respond in the Attack condition than the other groups. Surprisingly, no significant main effect was found for Group, $F(2,78) = 1.37$, $p = .26$, $\eta_{ges}^2 = .02$. There was also no significant interaction between Group and Action, $F(2,78) = 2.55$, $p = .084$, $\eta_{ges}^2 = 0.02$. However, we felt that direct pairwise comparisons to look at the effect of Group on the Attack condition of Action independently were justified, as this comparison was only indirectly tested by the main and interaction effect analyses described. Levene's test for homogeneity of variance approached significance ($F[2,78] = 3.09$, $p = .051$) so we opted to use independent *t*-tests without the assumption of equal variance for these pairwise comparisons. Table 5 shows the results. In summary, the AFOs were significantly faster than both novice groups when the virtual human attacked with either a Knife or Handgun. The effect size for both comparisons was large.

Table 5. Pairwise comparisons of the effect of Group on response times in the Attack condition of the Decision stage.

Group	*M* (ms)	*SD* (ms)	Bonferroni correct *p*-values / Cohen's *d*		
			AFOs	Matched Novices	Younger Novices
AFOs	507	42.2	-		
Matched Novices	565	65.8	.001* / 1.06	-	
Younger Novices	584	78.2	<.001* / 1.23	1 / 0.26	-

Note, * marks significance.

5.3.1.2.4. Effect of age

When designing this experiment, we assumed that controlling for age between groups would be essential for valid comparisons of behavioural and EEG measures. To test this assumption, we considered how much age explained variance within each group. Correlations between age and Preparation and Decision stage response times were calculated. The results of this analysis can be found in Table 6 and visualised in Figure 19. In summary, moderate to large positive correlations were observed in the AFO and Matched Novice groups, and no significant correlations were observed in the Younger Novices group.

For the AFO group we also looked at the relationship of their reported number of years' experience as an AFO, and the amount of time they had been a police officer, on response times in the experiment. Table 7 shows the correlation table of these individual measures with each other, and the response times. Unsurprisingly, the individual measures all correlated highly with each other. Further, their relationship with response times all showed the same general positive relationship. For the most part, the number of years as a police officer did not significantly correlate with response time and, where

it did, the other two individual methods better explained the variance. No clear differences were observed between the amount of variance explained by age versus number of years of AFO experience.

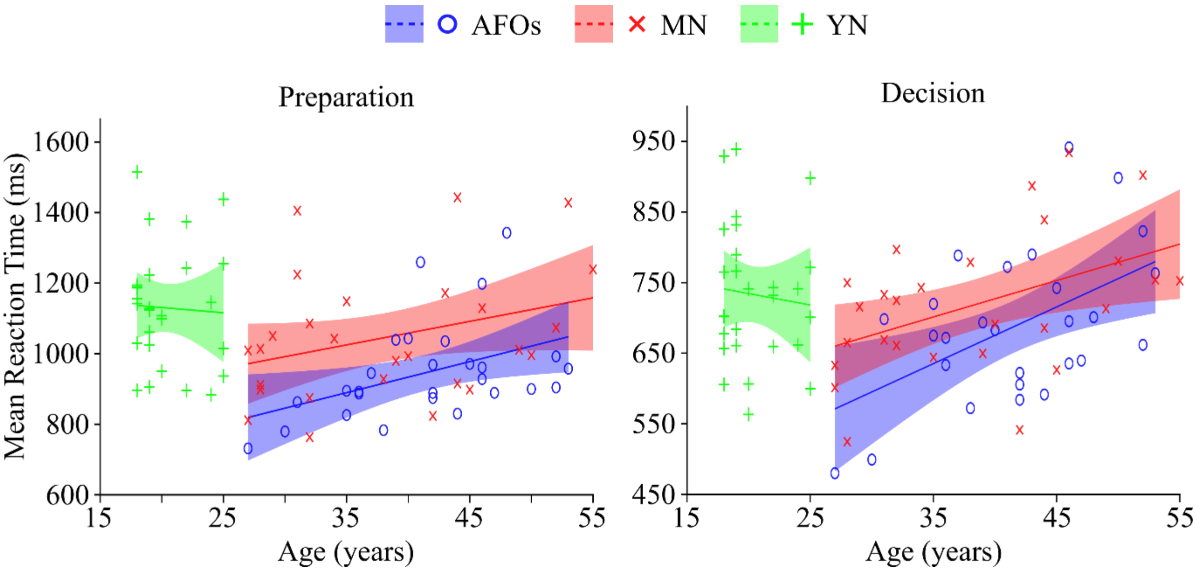


Figure 19. Findings from our analysis of the effects of age on response time at both stages of the experiment. Positive correlations between age and response time can be seen in the AFO and Matched Novice groups. This figure was originally produced in colour. Younger Novices (YN) data are marked by a green plus (+); Matched Novices (MN) data are marked by a red cross (x); AFOs data are marked by a blue oval (O). Not shown are correlations calculated for combined novice groups. For those tests, and statistics from all correlations, refer to Table 6.

For the AFO group we also looked at the relationship of their reported number of years' experience as an AFO, and the amount of time they had been a police officer, on response times in the experiment. Table 7 shows the correlation table of these individual measures with each other, and the response times. Unsurprisingly, the individual measures all correlated highly with each other. Further, their relationship with response times all showed the same general positive relationship. For the most part, the number of years as a police officer did not significantly correlate with response time and, where it did, the other two individual methods better explained the variance. No clear differences were observed between the amount of variance explained by age versus number of years of AFO experience.

Table 6. Summary of results from correlation analysis of the effects of age on response time. Results are shown for response times at both the Preparation and Decision stages of the experiment for each group individually. Additionally, results from a combined novice group are shown. Pearson's linear correlation was used to produce the correlation coefficient (*r*) and the significance test value (*p*).

Group	*df*	Experiment Stage			
		Preparation		Decision	
		r	*p*	*r*	*p*
AFOs	25	.43	.025*	.51	.006*
Matched Novices	25	.33	.096	.46	.016*
Younger Novices	25	-.05	.82	-.09	.67
All Novices	52	.06	.645	.11	.42

Note, * marks significance and *df* = degrees of freedom.

Table 7. Summary of results from Pearson's linear correlation analysis of the effects of AFO expertise on response time. Results for two separate measures of expertise, AFO Years and Police Officer Years, are shown. The values given in this table are the correlation coefficient (*r*) of each analysis.

Independent Variables		Individual measures		
		Age	AFO Years	Police Officer Years
Individual measures	AFO Years	.83***	-	-
	Police Officer Years	.85***	.82***	-
Preparation	Equip Nothing	.31	.35	.29
	Equip Taser	.40*	.44*	.36
	Equip Glock	.20	.18	.24
Decision	Don't Shoot Taser	.43*	.48*	.38
	Don't Shoot Glock	.42*	.46*	.41*
	Shoot Taser	.54**	.43*	.41*
	Shoot Glock	.39*	.25	.26

Note, *p*-values are marked as follows: *** $p < .001$, ** $p < .01$, * $p < .05$

5.3.2.1.5. Effect of sex

As the AFO and, by necessity, the Matched Novice groups only had one female participant each, it was not plausible to evaluate the effects of sex on response times within these groups. However, the Younger Novices group was relatively evenly split between men and women, so exploratory comparisons could be made. Just as for the previous experiments, the data for the Younger Novices group was refactored so that sex was a between-subject independent variable. See Figure 20 for an overview of the results.

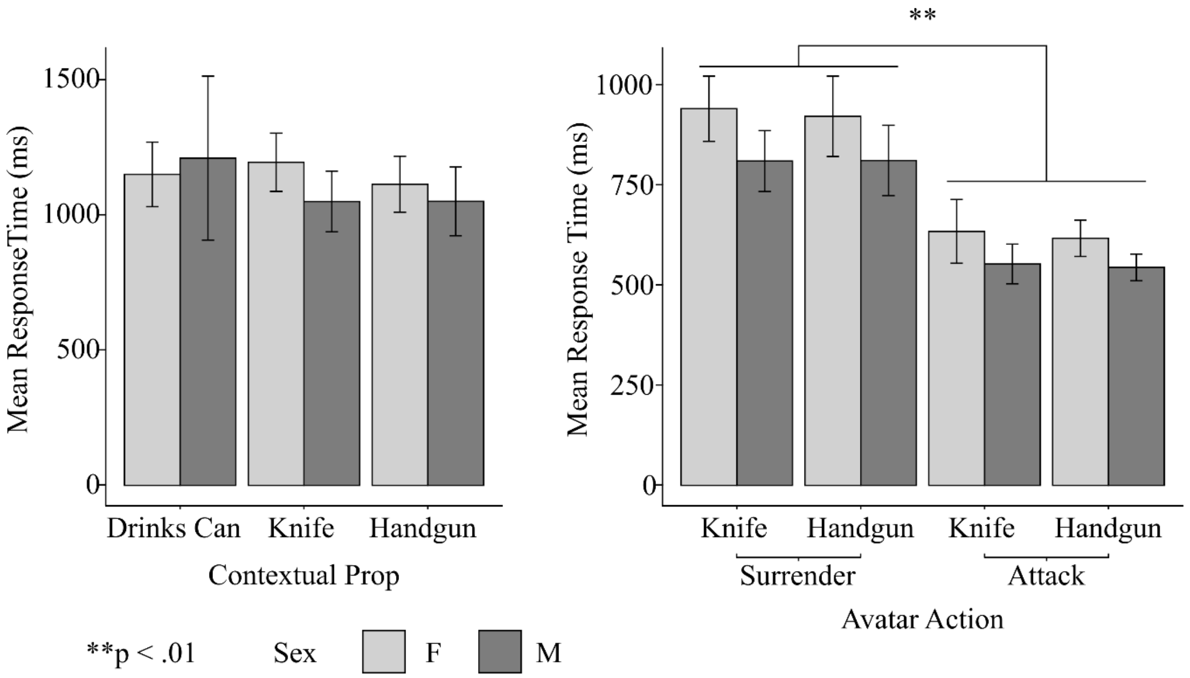


Figure 20. Effects of sex on response time within the Younger Novices group. Men were found to be significantly faster at the decision stage only. The left figure shows results from the Preparation stage and the right figure from the Decision stage. Significant findings from the analysis of variance are marked by connecting lines.

For the Preparation dataset, a two (between-subject, sex: men vs. women) by three (within-subject, Contextual Prop: Can vs. Knife vs. Handgun) mixed factor analysis of variance was conducted. No significant main effect of sex was observed, $F(1,25) = .55$, $p = .46$, $\eta_{ges}^2 = .01$. Mauchly's test suggested that the assumption of sphericity had been violated for the Contextual Prop factor, so degrees of freedom were corrected for using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.61$). No significant interaction between sex and Contextual Prop was found, $F(1.22,30.5) = 8.53$, $p = .275$, $\eta_{ges}^2 = .03$. Not reported are the main effect of Contextual Prop, as these formed part of the all group analysis.

A two (between-subject, sex: men vs. women) by two (within-subject, Contextual Prop: Knife vs. Handgun) by two (within-subject, Action: Surrender vs. Attack) mixed factor analysis of variance was conducted on the Decision dataset. A significant main effect of sex was observed, $F(1,25) = 8.53$, $p = .007$, $\eta_{ges}^2 = .14$. No significant interactions between sex and the other factors were found. This suggests that men ($M = 678$, $SD = 64.9$) were faster than women ($M = 777$, $SD = 101.3$) at the Decision stage, in the Younger Novices group. As before, the effects and interactions of Contextual Prop and Action have already been reported in the all group analysis.

5.3.3. EEG results

5.3.3.1 Pre-processing of EEG

We used the same pre-processing pipeline developed using the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld et al., 2011) for the earlier experiments, but with some improvements and adaptations to suit the current data better. I will summarise the established pipeline and highlight these changes.

Rather than remove M1 and M2 electrodes after recording, we simply did not apply gel to them. The median impedance value of the remaining channels was higher than previous experiments (10.2 kΩ), but still acceptable (Ferree et al., 2001). Electrodes were re-referenced from CPz to the common average. We used the same bandpass filter, but with a wider 1-35Hz setting. The lower high-pass component was chosen to allow analysis of lower frequencies in the longer trial epochs.

The continuous data was broken up into individual trials defined from three seconds before the Contextual Prop stimulus presentation and three seconds after the participants' final response or Action–whichever came later. Incorrect responses, 'no action' trials and trials where the participant missed their shot were removed before cleaning the data. Artefacts were detected and removed using the 'pre-cleaning' method described earlier. Briefly, persistent artefacts (e.g. eye blinks) were isolated and removed using independent components analysis, and trials still containing artefacts were removed entirely. No channels were removed from any dataset.

5.3.3.2. Time-frequency analysis

Before running the time-frequency analysis, trials were broken into smaller epochs around each event. These were from three seconds before to three seconds after each event in a trial. The original, full-length epochs were maintained for visualisation, as will be seen later.

The same time-frequency analysis pipeline was used as in the previous experiments but including the wider 1-35Hz range, with a spectral resolution of 1Hz. In short, we used Fieldtrip's multi-taper method convolution with a single Hanning taper and sliding windows in steps of 50ms. Time windows were set to four times the wavelength of the frequency analysed.

Trials and their epochs were then divided into subsets of conditions before baseline conversion. We used a baseline period from two to one second before the Preparation stage began. All baseline periods for subsets of trials were averaged before being used to convert activity periods into decibel values. The chosen baseline period was appropriate because participants were standing waiting for the virtual human to appear at the entranceway and should not have been moving at all. Panel A of Figure 21 shows the raw time-frequency data over the duration of a trial. Importantly, there is little variation in activity during this period. Events can be clearly seen as changes in power relative to baseline across multiple frequencies. Panel B of Figure 21 shows variation between subsets of trials. These provide context for interpretation of results in the next section.

5.3.3.3. Introduction to results

The mixed factor design of this experiment means that there were many possible comparisons that could be made. However, unlike for the behavioural data, we could only make one comparison between EEG datasets at a time without reducing the data to a single dimension. In the earlier experiments, we only had two conditions, so this was not an issue, but the introduction of between-subject factors and the more complex design of the experiment meant that more careful planning was required.

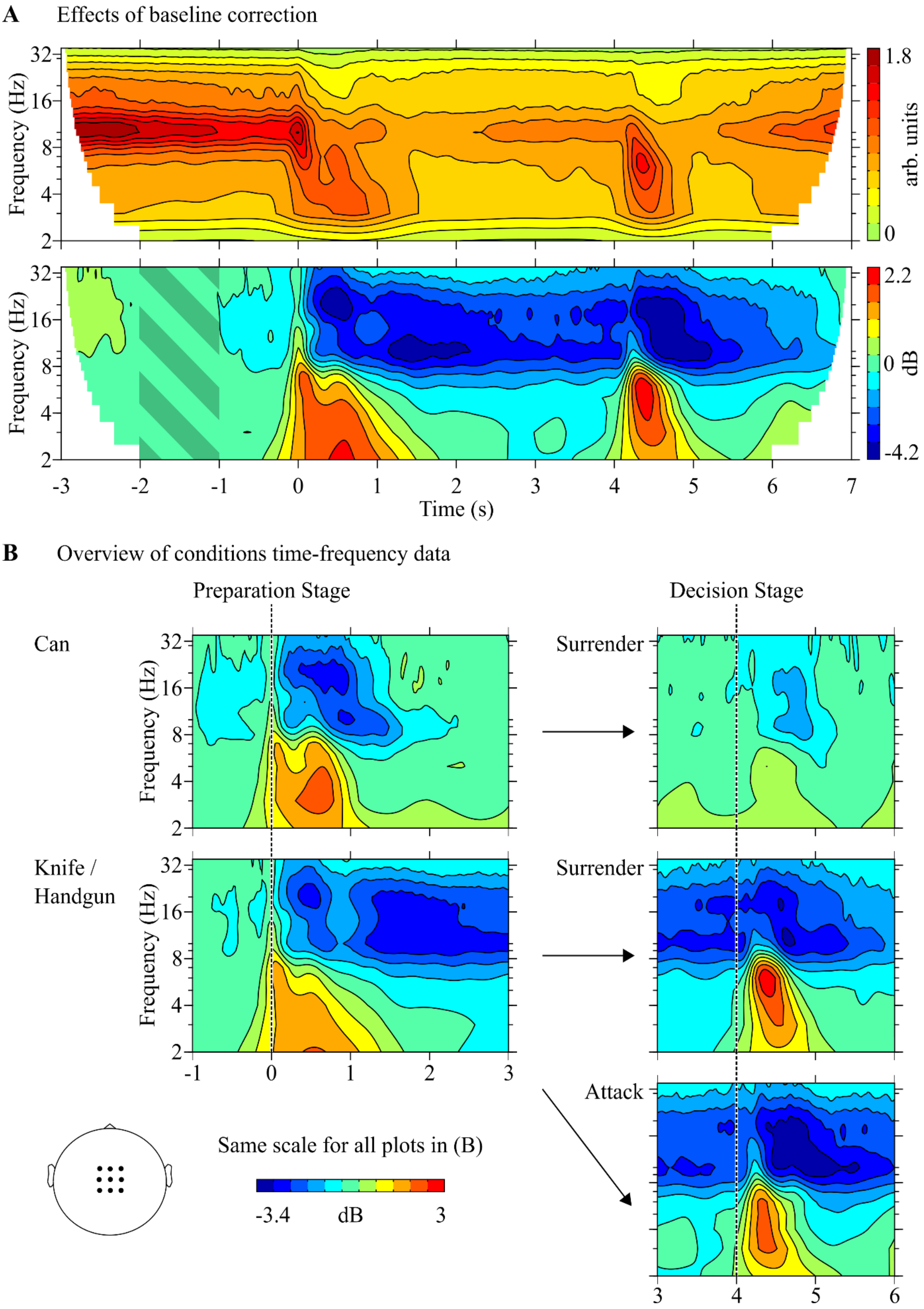


Figure 21. Overview of time-frequency data in the complex scenarios experiment. The overall trend and variation between conditions can be seen as a guide for our subsequent analysis. Data for the AFO group is shown as representative of the general effect seen in all groups. All plots are based on

the central 9 electrodes, highlighted in the bottom left. (A) First shows the raw relative power across all conditions (only correct responses). The purpose of this is to evidence the choice of baseline period between -2 and 1 second before the first prop was presented. The lower panel shows the same data after being compared to this baseline using decibel conversion. (B) Shows both the experimental design and its effect on the recorded EEG in the time-frequency domain.

I have divided the presentation of results into sections containing within-subject comparisons and between-subject comparisons. In both cases we were primarily interested in pre-response locked activity at the Preparation and Decision stages. For within-subject comparisons, all groups underwent the same analysis. Comparisons that would reveal differences between threatening and non-threatening conditions and between different levels of threat were of interest. For between-subject comparisons we were only concerned with comparing AFOs with the Matched Novices. Data from the Younger Novices is given only to supplement those comparisons. This was because we only wanted to compare experts with novices and were not specifically interested in the effects of age and sex.

Analysis was primarily conducted at the sensor level using cluster-based analysis (Maris & Oostenveld, 2007). This was supplemented by source level analysis. The methods were almost identical to those used in the previous experiments. For sensor level analysis 25,000 permutations were used to create the Monte Carlo distributions which estimated how likely the two distributions being compared were equal. 10,000 permutations were used for source level analysis to save on computation time. For within-subject comparisons, the dependent *t*-test ($df = 26$, $|t| > 2.05$) was used as the test statistic in cluster-based analysis. For between-subject comparisons, the independent *t*-test ($df = 52$, $|t| > 2$) was used.

5.3.2.4. Within-subject EEG results

5.3.2.4.1. Equip firearm vs. equip nothing

At the Preparation stage participants could either equip a firearm (Glock or Taser) or nothing (Safety). They did so in response to identifying the Contextual Prop held in the virtual humans' hand. Differences in response-locked EEG needed to be carefully interpreted as participants' actions varied greatly depending on the response. However, in the pre-response period, defined as 600ms to 0ms before response, all stimuli across all conditions were comparable. A 600ms window was chosen because participant response time was rarely fast enough that the stimulus had not been presented at that point. The longer time period was useful for estimating non-transient changes in lower frequencies. Results from sensor level analysis can be seen in Figure 22.

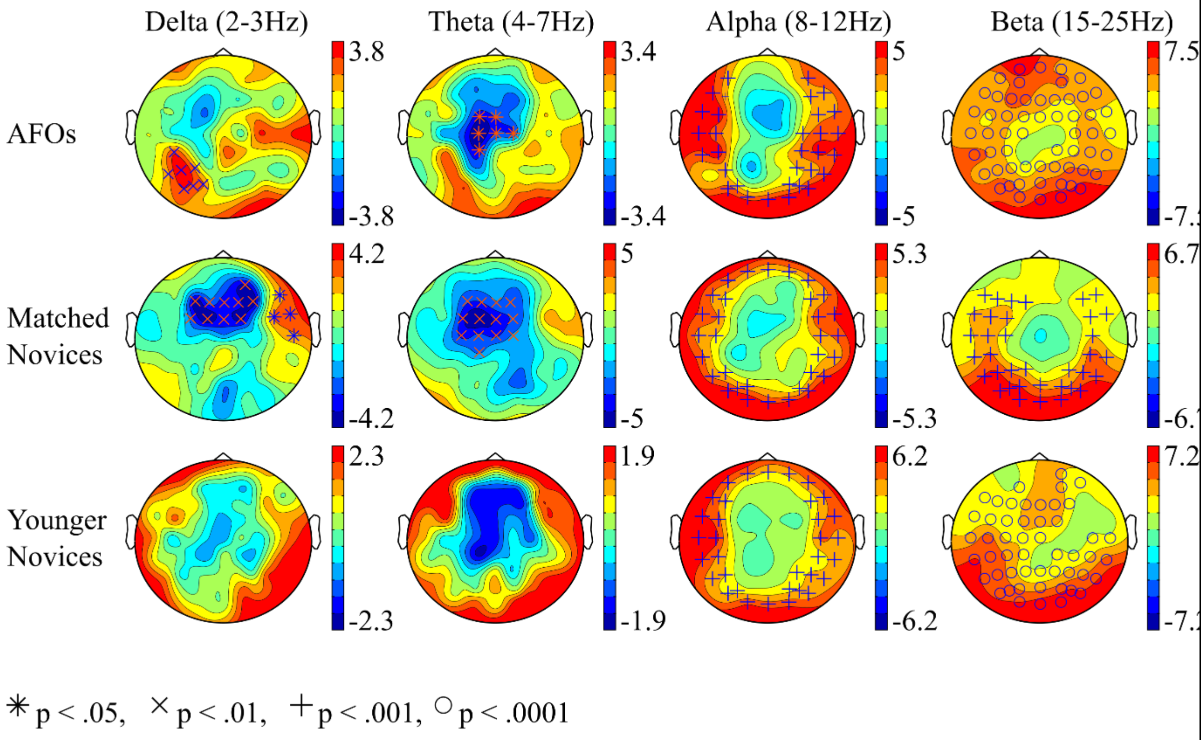


Figure 22. The differences between EEG recorded before either a firearm was equipped (Taser or Glock) or when nothing was equipped (i.e. Safety button pressed). Notably, all groups show the expected pattern of greater frontal-midline theta activity in the Surrender condition. Each topography shows the difference between these two conditions (firearm minus nothing) in the pre-response period from -600ms to 0ms. Comparisons for all frequency bands of interest by all groups are shown. Significant clusters are marked over included electrodes. All scales represent *t*-values from the cluster analyses.

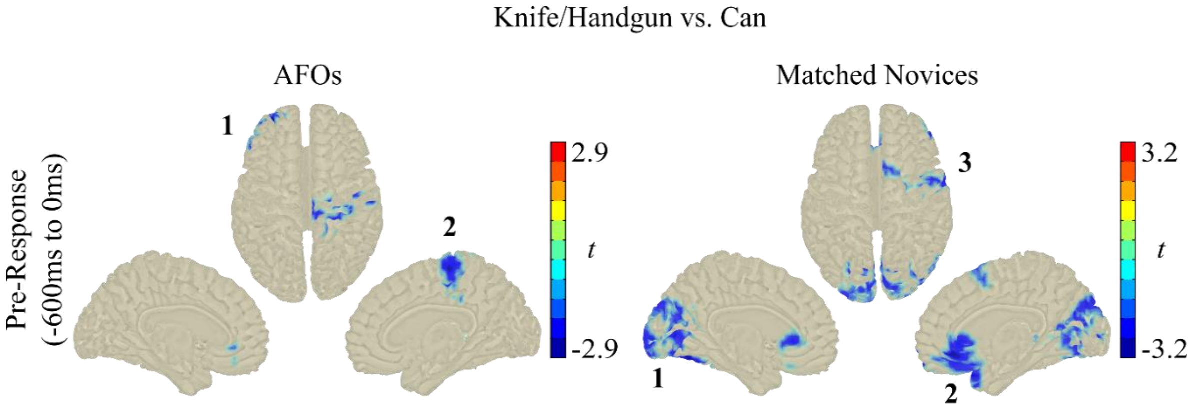
Our main prediction was that this comparison would have similar results to the Contextual Prop experiment, as it involved a similar task: identifying and responding to threat. In particular, we were expecting greater theta activity in the Can condition than in the Knife or Handgun conditions. For the AFOs, one significant negative cluster, $p = .029$, 95% CI [.027 .031], was found over central electrodes. A positive cluster was also found over the left parietal electrodes, but was not significant, $p = .074$, 95%

CI [.071 .077]. A similar significant negative cluster, $p = .006$, 95% CI [.005 .007], was found running the same test on the Matched Novice group, over left fronto-central electrodes. While the topography of differences was similar in the Younger Novices group, the effect was considerably smaller, and no clusters were identified for testing.

We were not able to measure oscillatory activity in frequencies lower than theta in the earlier experiments so our predictions for the current experiment were more limited. However, the role of delta in inhibition of action (Harmony, 2013) meant it was an interesting target for comparison both within and between groups. Delta is variably defined within the 0.5 to 4Hz range. For resting state with very long epochs it might be feasible to measure lower than 1Hz, but for our purposes we defined delta as 2-3Hz. In the AFO group, two positive clusters were found. One was significant and closely matched the positive theta cluster over the left parietal electrodes, $p = .009$, 95% CI [.008 .011]. The other was not significant, $p = .058$, 95% CI [.055 .061]. Taken in consideration with the differences in theta, two separate sources of activity emerge in the AFO group: one fronto-central and the other left parietal. However, in the Matched Novices group, delta activity was more closely aligned with theta activity: one significant negative cluster, $p = .004$, 95% CI [.003 .005], was found over fronto-central electrodes. A significant positive cluster was also found in right fronto-temporal electrodes, $p = .04$, 95% CI [.037 .042], also matching the topography of theta activity. The Younger Novice differences in delta matched the differences in theta in that no clusters were identified. The topographies were broadly the same in both frequency bands.

Differences in pre-response alpha and beta activity were similar in all groups. A single significant positive cluster ($p < .001$) was found for all comparisons. Alpha power decrease was greater in the Can condition than the Knife/Handgun conditions across occipital and temporal electrodes. There were no differences in alpha activity over central electrodes. Decreases in beta activity were greater over occipital and parietal electrodes for all groups, and fronto-polar electrodes for the AFO group. Note, post-response (0 to 600ms) beta activity showed the expected greater decrease over central electrodes in the trials where participants moved considerably more having equipped their firearm, so this effect was specific to pre-response activity.

We had planned to run the same analysis of theta band activity at source level, using the methods described in the previous experiments. Briefly, we used the structured light scans to create individual head models and electrode positions which were input into an analysis pipeline based on eLORETA (Pascual-Marqui et al., 1994) based analysis described in Chapter 4. Figure 23 shows the output of this analysis. Clusters have been numbered in the figure for reference. Note, two only borderline significant findings have been included as the confidence interval of Monte Carlo estimation included our threshold of significance.



1: $p = .052$, 95% CI [.048 .056]
2: $p = .035$, 95% CI [.031 .039]

167 negative clusters, 2 positive clusters

1: $p = .041$, 95% CI [.037 .045]
2: $p = .021$, 95% CI [.019 .024]
3: $p = .051$, 95% CI [.047 .055]

184 negative clusters, 0 positive clusters

Figure 23. Source analysis results from the difference between pre-response (-600ms to 0ms) theta when participants equipped a firearm versus when they pressed Safety. In both groups, differences in the ACC can be seen. Only significant and borderline significant clusters are shown. These are labelled by proximity and results from the cluster-based analysis are shown below each plot.

For the AFOs, the analysis identified 167 negative clusters and two positive clusters. Two of the negative clusters were significant and suggested greater theta activity in the Can condition: the first in the left middle frontal region, $p = .052$, 95% CI [.048 .056], and the second in the right superior motor area, $p = .035$, 95% CI [.031 .039]. In the Matched Novice group, 184 negative clusters were identified and three were significant: the first covering the occipital lobe, $p = .041$, 95% CI [.037 .045]; the second in the right anterior cingulate cortex, $p = .021$, 95% CI [.019 .024]; the third in the right middle frontal region, $p = .051$, 95% CI [.047 .055].

5.3.2.4.2. Identifying different levels of threat

One of the biggest challenges of using naturalistic stimuli is creating conditions which vary only along the parameter you are interested in. In this experiment, the closest match of conditions was between the Knife and Handgun conditions of the Contextual Prop independent variable. The required response only varied in which firearm was equipped, Glock or Taser. We were interested to see if the same cognitive functions used to determine threat versus no threat applied to the distinction between higher and lower levels of threat. To do this, we used the same data from the Preparation stage, considering only pre-response (-600ms to 0ms) changes in oscillatory power. The results of all comparisons can be seen in Figure 24.

In the AFO group, delta, theta and beta oscillatory bands all had significant positive clusters, comparing Handgun with Knife. In delta there were two significant positive clusters over the right frontal, $p = .002$, 95% CI [.001 .002], and occipital electrodes, $p = .002$, 95% CI [.002 .003]. A single cluster in theta, $p < .001$, 95% CI [<.001 <.001], was found over both of these areas. Results from delta and theta suggest greater increase in activity when identifying and responding to the Knife rather than the Handgun. In beta the significant positive cluster, $p = .013$, 95% CI [.012 .015], suggested greater decrease in beta for the Knife condition over occipital and right centro-parietal electrodes. A similar, but not significant, cluster was found in alpha band, $p = .06$, 95% CI [.057 .063].

The Matched Novice and Younger Novice groups did not have as widespread a difference in pre-response EEG activity between high and low threat conditions. The only difference for the Matched Novice group was in theta band. A significant positive cluster, $p = .001$, 95% CI [.001 .001], was found over parieto-occipital electrodes, matching the pattern seen in the AFO group. In the Younger Novice group, a significant positive cluster was found for delta over right fronto-temporal electrodes, $p = .034$, 95% CI [.032 .036], and beta over left parietal electrodes, $p = .036$, 95% CI [.033 .038].

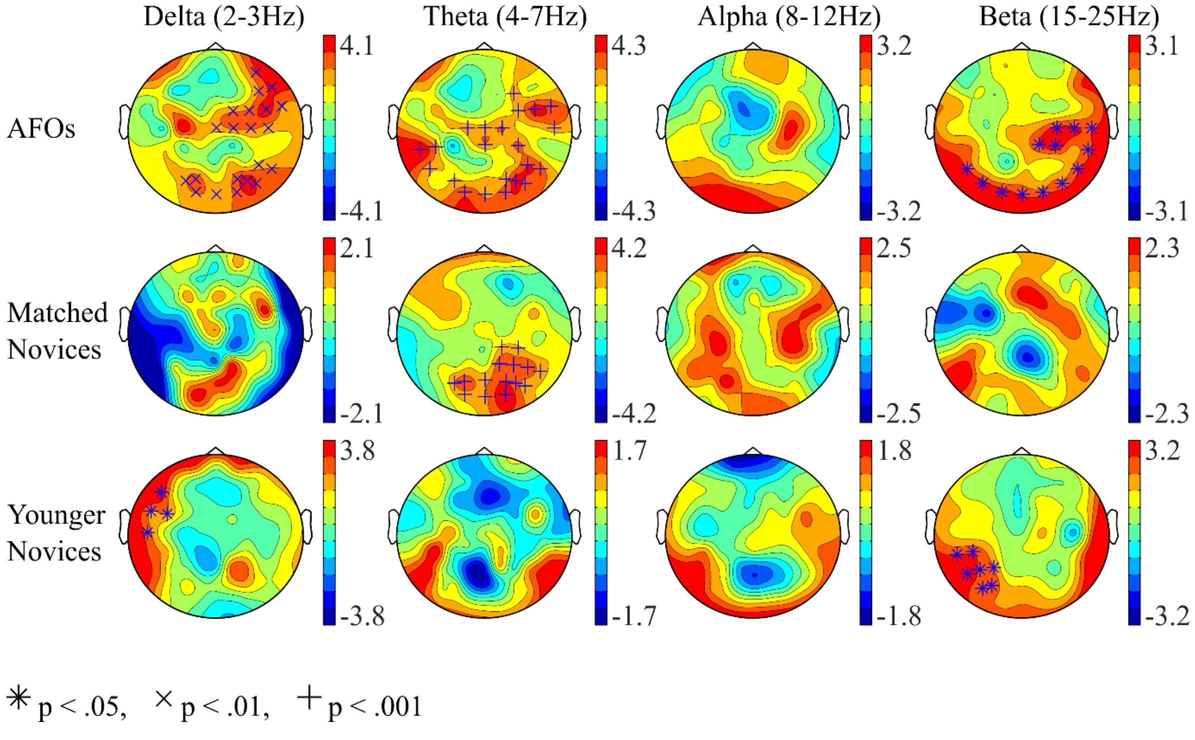


Figure 24. The differences in pre-response EEG (-600ms to 0ms) recorded before either a Glock or a Taser was equipped at the Preparation stage. Generally, AFOs showed greater differences between these conditions. Most notably, they had greater central delta in the Glock condition. Comparisons for all frequency bands of interest and for all groups are shown. All scales are *t*-values from the relevant cluster analyses (Glock minus Taser). Significant clusters are marked over electrode positions.

5.3.2.4.3. Shoot versus Safety

The critical moment in all threatening trials was the decision to shoot or press the Safety. This decision is the focus to our 'shoot/don't-shoot' paradigm. In this task, the animation of the virtual human determined whether the participant should shoot or not, just as for the Action experiment. Therefore, we predicted similar patterns of changes in oscillatory power in the current experiment. Specifically, a shift so that the Attack condition would have a greater increase in central theta activity than the Surrender condition. To test for differences, we considered the pre-response (-400ms to 0ms) time-frequency data at the Decision stage when a Knife or a Handgun had been correctly responded to at the earlier Preparation stage. This time window was chosen because of the faster reaction times in the Attack condition. The results of these comparisons within all three groups can be seen in Figure 25.

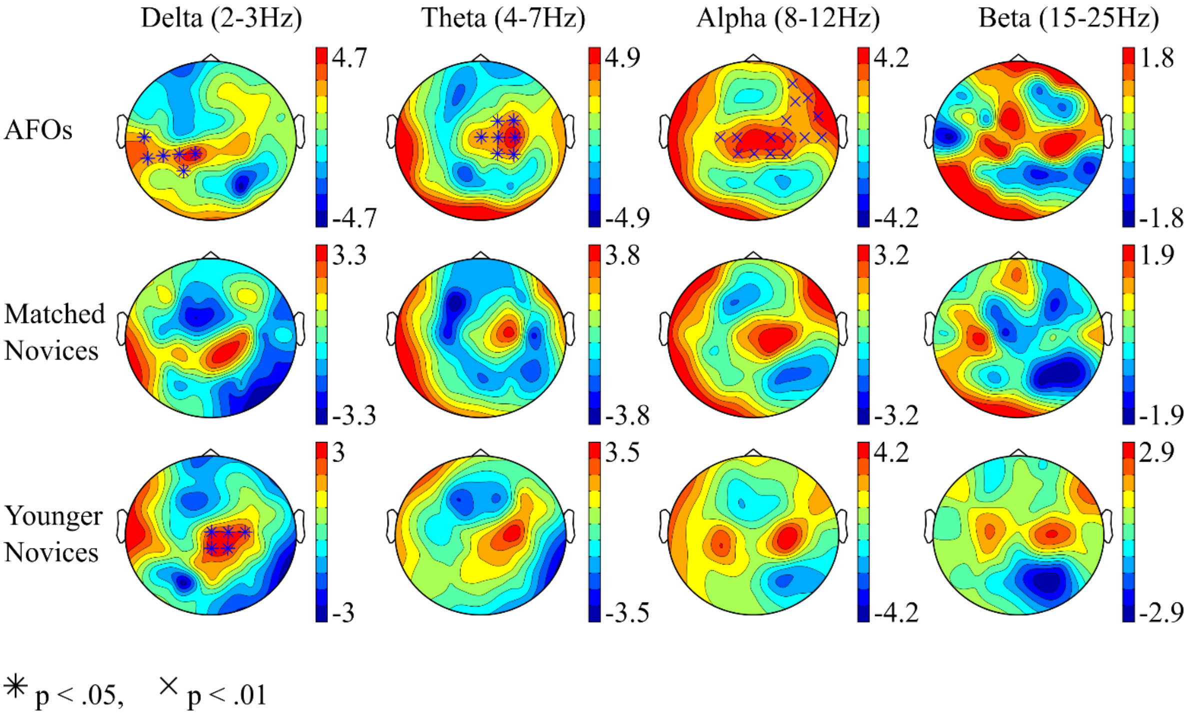


Figure 25. The differences in pre-response EEG recorded before either a firearm (Taser or Glock) was discharged or when the Safety button was pressed while holding a firearm. Topographies were consistent with our predictions, based on our earlier Contextual Prop experiment. Each topography shows the difference between these two conditions (shoot vs press Safety) in the pre-response period, from -600ms to 0ms. Comparisons for all frequency bands of interest by all groups are shown. Significant clusters are marked over included electrodes. All scales represent *t*-values from the cluster analyses.

Viewing the results from the three groups for delta, theta and alpha frequencies, a clear pattern emerges. Despite the differences between groups and noise introduced from four epochs (baseline and activity of both conditions), the topographies in these lower frequencies are remarkably similar. From 2-12 Hz, all groups show a greater power increase in the Attack condition over central electrodes and greater increase in the Surrender condition over surrounding frontal and parietal electrodes. In each case, either positive clusters, negative clusters, or both, were identified. However, these topographies also show a limitation of cluster-based analysis; it is not sensitive to changes (however large) in small numbers of electrodes when neighbouring clusters have the opposite effect.

To determine if these clusters were the result of two nearby sources or (more likely) a single dipolar source, we used source localisation. Figure 26 shows the results of this analysis. We compared pre-response (-400ms to 0ms) theta (3-7Hz; wider band for source analysis) in the AFO and Matched Novices groups. For the AFOs, 114 positive clusters were identified, one of which was significant, p = .036, 95% CI [.034 .038], centred around left pre-central gyrus. In the Matched Novices group, 13 positive and 38 negative clusters were identified, but none were significant. However, a similar pattern of activity as the AFOs can be seen. These results align well with the sensor level analysis, possibly suggesting similar sources which differ in magnitude.

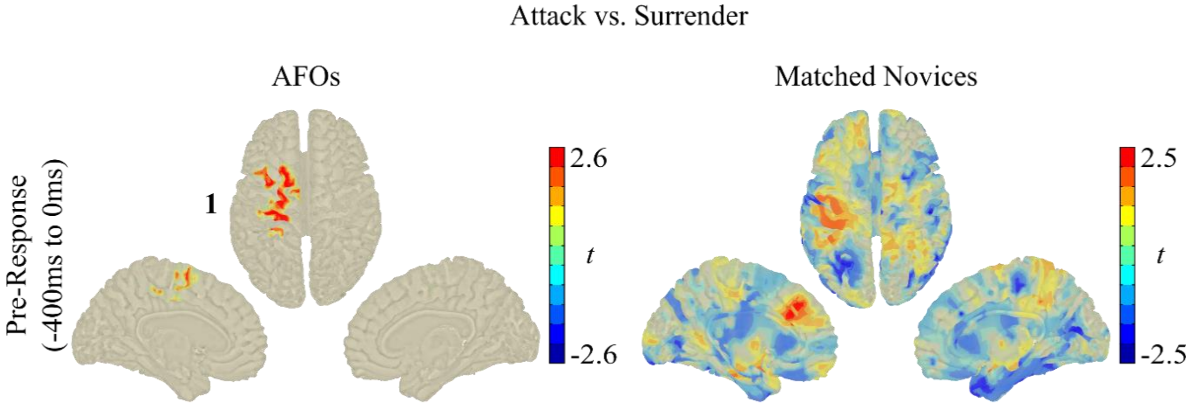


Figure 26. Source analysis results from the difference between pre-response (-400ms to 0ms) theta when participants shoot or press Safety. While consistent with our sensor level analysis (Figure 25), our findings were not conclusive. In the AFO group, only the significant cluster is shown. No significant clusters were identified in the Matched Novices group, so the overall results from the comparison are shown.

5.3.2.4.4. Exploratory analysis

We had no predictions about whether the type of firearm being fired would have any effect on pre-response EEG. However, we wanted to check whether the differences between Attack and Surrender conditions were dependent on any other factor. Generally, we can conclude that the effects we have described apply to both the use of the Taser and Glock.

5.3.2.5. Between-subject results

5.3.2.5.1. Introduction

The within-subject analyses suggested there may have been some differences between groups, as some differences between conditions were found to be significant in one group but not another. However, many conditions exist that allow one group to appear to have a greater effect than another, when no observable difference is present. For example, a test may be slightly above the threshold of significance and another may be slightly below. Therefore, we needed to make comparisons between groups when they were undergoing the same conditions. Figure 27 shows the differences in time-frequency data for each condition between the AFO and Matched Novice groups and should be used as a guide when interpreting cluster-based analysis results.

The importance of baseline correction has already been discussed, but it should be emphasised as a requirement for between-subject comparisons. The signal recorded at each electrode is proportional to the energy at the scalp, but many factors, such as impedance between scalp and electrode, influence how great this proportion is. This means that the unit of power calculated by time-frequency analysis is arbitrary and changes for each electrode while in use and between applications. For within-subject comparisons, a counterbalanced design ensures that any changes in this proportion over time are evened out between conditions and so their effect on statistics is negligible. However, for between-subject comparisons there may be systematic differences between individuals. Baseline conversion after time-frequency analysis creates a standard unit of power that can be compared between electrodes and individuals.

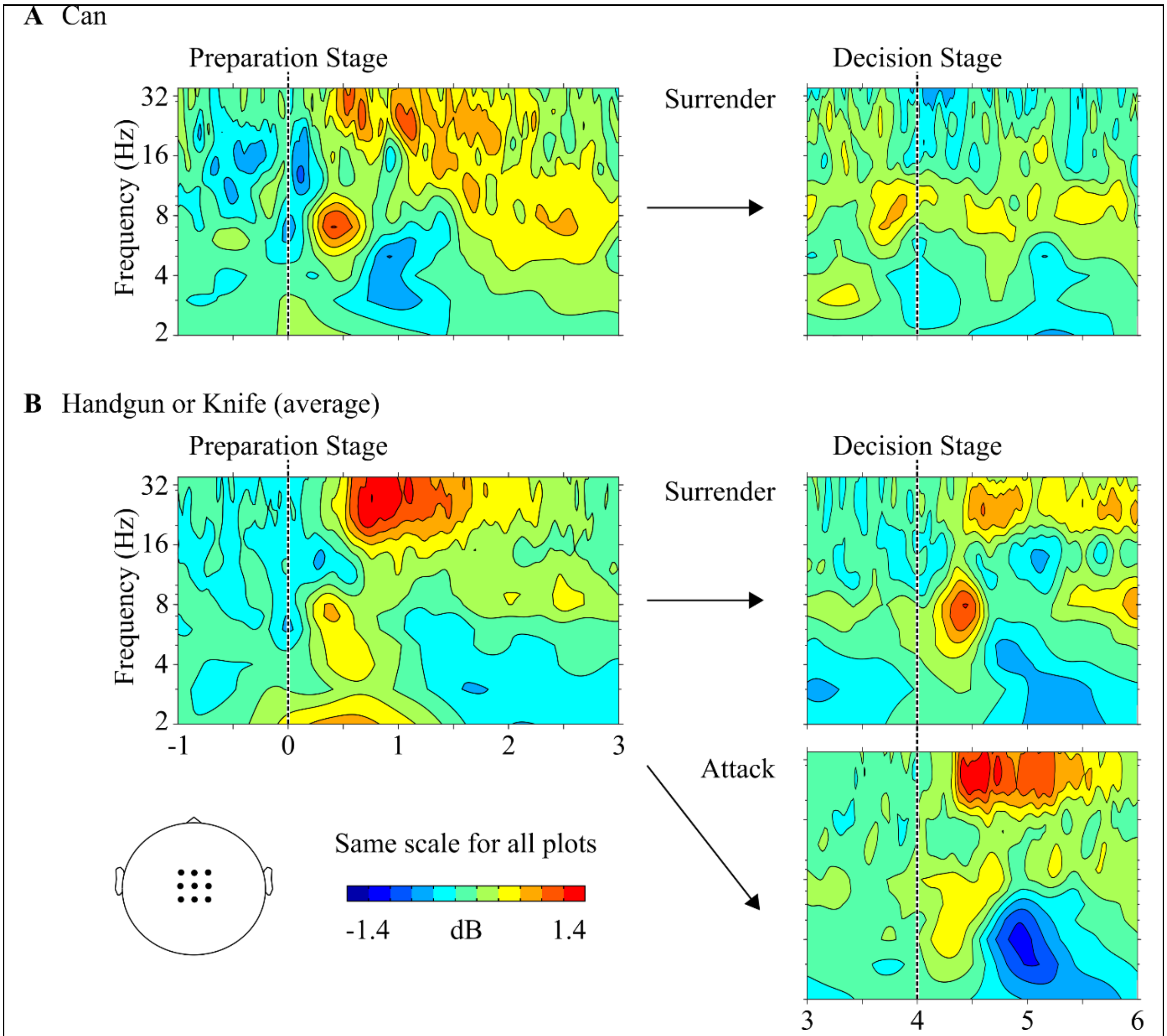


Figure 27. Time-frequency representation of the differences between AFO and Matched Novice groups (AFO minus Matched Novice). This figure provides an overview to aid interpretation of our cluster-based analysis. A) Shows the difference between groups when the Contextual Prop is a Can and always followed by the Action of Surrender. B) Shows the difference between groups when the Contextual Prop is threatening (Knife or Handgun), followed by either the Action of Surrender or Attack. All plots show stimulus-locked data from the central nine electrodes. Stimulus onset is marked by a dashed line. The scale is in dB calculated against the average baseline of all conditions.

5.3.2.5.2. Preparation stage

We compared pre-response (-600ms to 0ms) changes in delta, theta, alpha and beta band oscillatory power between groups in the Knife/Handgun and Can conditions of the Contextual Prop independent variable. A summary of these differences, as well as supplementary post-response differences, can be seen in Figure 28.

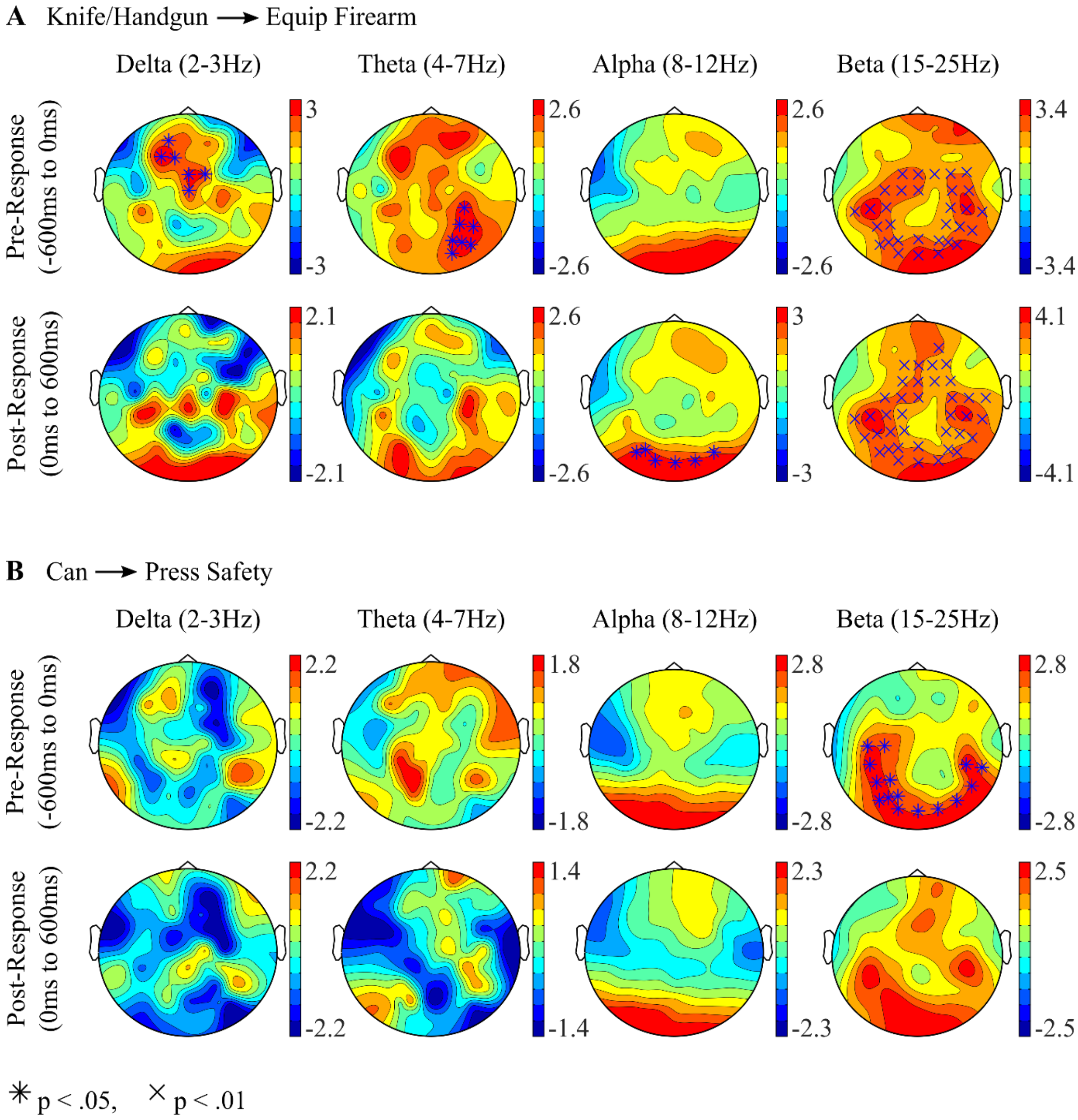


Figure 28. Between-subject differences in pre- and post-response EEG (-600ms to 0ms and 0 to 600ms, respectively) at the Preparation stage. Comparisons between the AFOs and Matched Novices are shown. This figure shows that the differences we observed between AFOs and Matched Novices at the Preparation stage were driven by between-subject differences in the threatening, Knife/Handgun conditions. (A) Shows the results of cluster-based analysis between the average of both firearms conditions (Glock and Taser). (B) Shows the results of cluster-based analysis for the press Safety condition. Significant clusters are marked, and all scales represent *t*-values from the relevant cluster analyses.

First, we compared the two groups equipping a firearm. For delta band, a significant positive cluster, $p = .023$, 95% CI [.027 .031], over frontal electrodes suggested greater increase in the AFO group versus the Matched Novices. A similar, but non-significant positive cluster, $p = .078$, 95% CI [.074 .081], was found in theta band. A larger, significant cluster, $p = .038$, 95% CI [.036 .041], suggested that the difference in theta was more posterior. However, these clusters may represent a single source across the two frequency bands. A positive cluster in alpha was found over occipital electrodes, but was not significant, $p = .056$, 95% CI [.054 .059]. A significant positive cluster in beta was found over eccentric centro-parietal and occipital electrodes, $p = .01$, 95% CI [.009 .011], suggesting greater decrease in beta in the Matched Novice group. This effect continued post-response.

We found similar differences in alpha and beta band in the Can condition. A positive cluster in alpha over occipital electrodes did not reach significance, $p = .065$, 95% CI [.062 .068], but showed a similar trend. A smaller, but comparable significant positive cluster, $p = .02$, 95% CI [.018 .022], in beta suggested greater decrease in beta activity over the eccentric centro-parietal and occipital electrodes in the Matched Novice group. No clusters were identified for testing in theta or delta oscillatory bands.

5.3.2.5.3. Decision stage

We ran more comparisons at the Decision stage. Our primary interest was whether we could predict expertise from pre-response activity in the Attack and Surrender conditions. Therefore, we collapsed trials across Handgun and Knife conditions, and divided them into Attack and Surrender for Action. The results of the comparisons can be seen in Figure 29.

No clusters were identified for testing between AFO and Matched Novice groups in the Attack condition for delta, theta or alpha band. However, a significant positive cluster, $p = .015$, 95% CI [.013 .016], suggested greater beta decrease in power in the Matched Novice group over occipital and temporo-parietal electrodes. A second, positive cluster was found over central electrodes, $p = .083$, 95% CI [.08 .086].

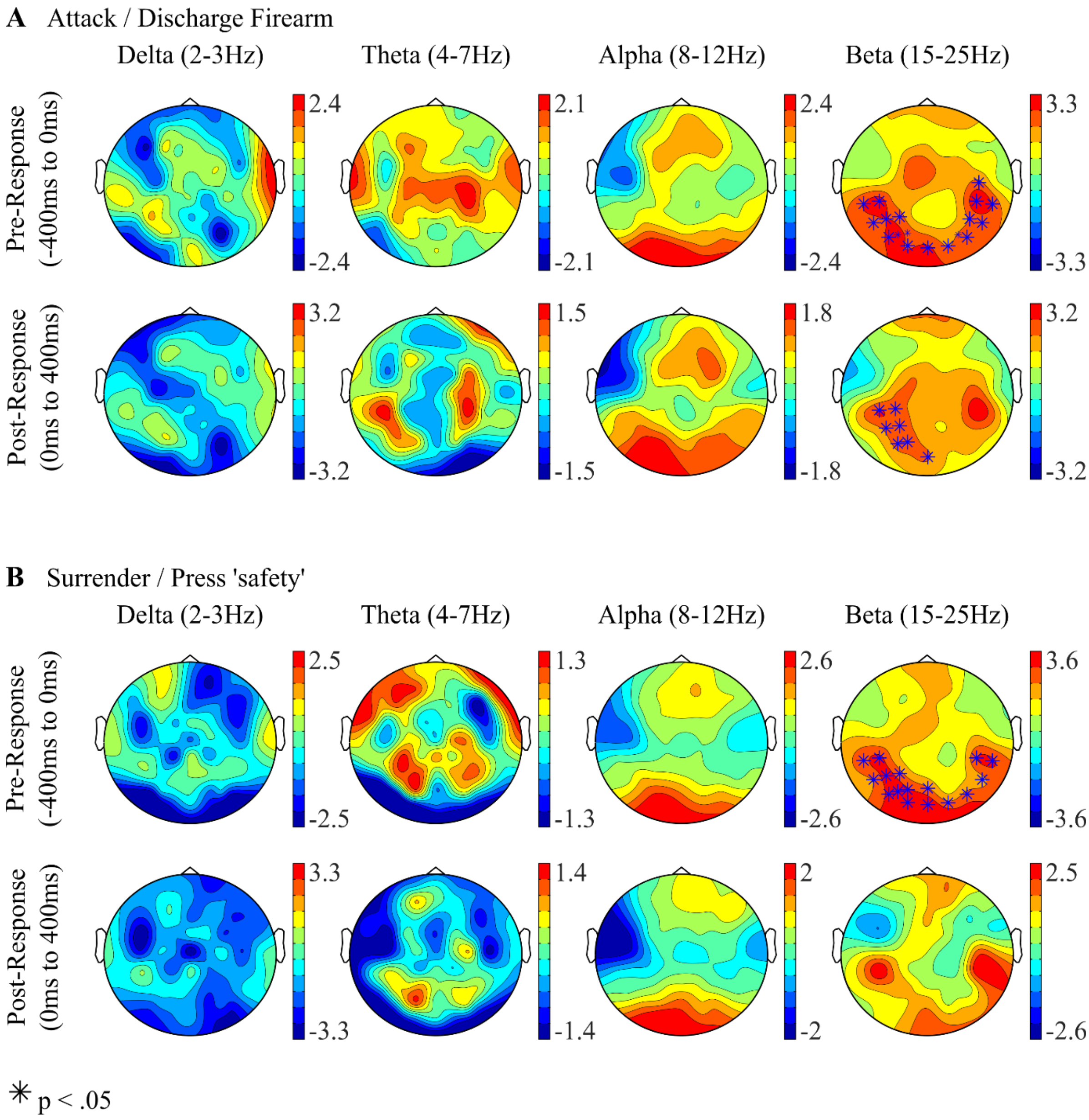


Figure 29. Between-subject differences in pre- and post-response EEG (-400ms to 0ms and 0 to 400ms, respectively) at the Decision stage. Comparisons between the AFOs and Matched Novices are shown. (A) Shows the results of cluster-based analysis between the average of both conditions where a firearm was discharged (Shoot Glock and Shoot Taser). (B) Shows the results of cluster-based analysis for the press Safety condition. Significant clusters are marked, and all scales represent *t*-values from the relevant cluster analyses.

Almost identical observations were made for the Surrender condition comparison between AFOs and Matched Novices. No clusters were identified for delta, theta or alpha band. A single significant positive cluster suggested greater beta decrease in the Matched Novice group over occipital and temporo-parietal electrodes, $p = .014$, 95% CI [.013 .015].

5.3.2.5.4. Identifying different levels of threat

To test between-subject differences in identification of different levels of threat, we first calculated within-subject differences between the Handgun and Knife conditions at the Preparation stage, as seen in Figure 24. Next, we compared the difference of these differences, between the AFO and Matched Novice groups. No clusters were identified for pre-response theta, alpha or beta band activity. However, we found two positive clusters in delta. The first was over central electrodes, $p = .026$, 95% CI [.022 .029] and the second over right occipital electrodes, $p = .027$, 95% CI [.022 .024]. These positive clusters suggest greater differences between the Handgun and Knife conditions in the AFO group than the Matched Novices group.

5.4. Interim discussion

5.4.1. Manipulation checks

The first, essential step in our analysis was to test whether AFOs performed better at the task than novice groups. As expected, participants across all groups made few errors, so performance was measured by response time. We found that AFOs had significantly faster response times than both Matched Novices and Younger Novices at the Preparation stage. This meant they were faster at identifying whether the virtual human was threatening and prepare themselves if he was. At the Decision stage of the experiment where participants had to decide whether to discharge their firearm, we found no significant main effect of Expertise, or any significant interaction with other independent variables. However, we found that AFOs were significantly faster to discharge their firearm than both novice groups. Note, no significant differences were found between the two novice groups at either the Preparation or Decision stages of the experiment, suggesting the cause of this differences was the expertise AFOs possessed. Overall, these findings suggest that the scenarios and related measures were sensitive to our expertise manipulation. This justified further analysis of the behavioural and electrophysiological data.

Another assumption made in the design of the experiment was that controlling for age and sex between groups would be essential to avoid introducing confounds alongside our expertise manipulation. Observed differences in response time between groups generally supported this assumption as Younger Novices were consistently slower than Matched Novices at both stages of the experiment and across all conditions. This shows that had we used a control sample more typical of a psychology/neuroscience study (i.e. the Younger Novices), the differences in reaction times we observed would have been exaggerated. However, as noted, no significant differences were found between the novice groups directly. In our exploratory analysis, we found significant positive correlations between age and response time in the Matched Novice and AFO groups, but not in the Younger Novice group, or when data from both novice groups were combined. Regarding sex differences, in the Younger Novice group, we found some differences between men and women at the Decision stage, but not the Preparation stage. Taken together, these findings do not provide evidence that age or sex affect the response times of novice participants. However, our data does suggest that, for AFOs, age and/or experience are important factors for performance. This will be discussed in more detail in the next section.

5.4.2. Summary of findings

5.4.2.1. Behavioural findings summary

Generally, within-subject comparisons of response times fit with our predictions. At the Preparation stage, we found a main effect of Contextual Prop which suggested that participants were faster to equip a Glock than they were to equip a Taser or indicate the virtual human was safe. While we did not find a difference in response times to equip a Taser versus indicate the virtual human was safe, this may be explained by differences in how participants gave their responses. To equip a Taser, participants first had to move their hand to their chest and then press a button. However, to indicate the virtual human was safe, participants only had to press a button which their thumb was already covering. It is likely that if some movement were required to indicate safety then response times would increase. Nonetheless, comparisons between this condition and the threatening conditions are not well controlled and should not be interpreted using behavioural data. Therefore, the finding of most interest was that participants were slower to equip a Taser than a Glock. One, or a combination of two, related factors could explain this difference: the position of participant's hands before the Preparation stage may have been biased towards equipping the Glock; and/or they may have been quicker to identify and respond to the higher level of threat posed by a Handgun versus a Knife. In either case, differences in response time suggest that participants were more oriented towards a higher level of threat. However, there was a third factor that may have explained differences in response time between these conditions. The position of the Glock on the hip may have made it easier to equip than the Taser, which was equipped on the chest. If experimental design was our only consideration, we would have introduced some counterbalancing measures for holster position. However, AFOs generally equip their Taser in a chest holster and the Glock on their hip or upper thigh, so, on balance, we decided to maintain realism and avoid introducing error from AFOs having to unlearn their trained behaviour. Even so, the position of holsters remains a caveat in this analysis.

At the Decision stage, we found that participants were significantly faster to discharge their firearm than they were to press Safety. This fit with our predictions based on our earlier studies, as well as findings reported elsewhere (e.g. Correll et al., 2006; Landman et al., 2016a). Generally, both the average response times and difference between conditions was similar to those reported for the Action experiment. Generally, participants responded to threat within the 1s animation time for the Action, meaning they shot the virtual human before he fully raised his arm (stabbing with Knife/aiming Handgun). Because of the introduction of the Preparation stage in the current experiment, we were able to compare response times when participants were using a Glock versus using a Taser. This comparison revealed no difference in response times for either action.

As discussed in the previous section, between-subject differences showed that AFOs had significantly faster response times at the Preparation stage and when discharging their firearm at the Decision stage. These findings show that AFOs performed better when responding to threat than novices. Only two responses in the experiment did not involve responding to threat: first, not equipping a firearm at the Preparation stage; and second, pressing Safety at the Decision stage. For both responses no difference between groups was found and response times were generally slow and variable. One possible reason for this is that, aside from following the task instructions, there was no motivation for participants to respond quickly in those conditions. Unlike in the threatening scenarios, there was no urgency for participants to protect themselves when the virtual human did not have a weapon or was surrendering. They could respond in their own time without consequence.

5.4.2.2. EEG findings summary

For the EEG analysis, we focused on comparing the main groups of interest, AFOs and Matched Novices, although data from the Younger Novices was still analysed. Our analysis was guided by the results of earlier experiments. We expected differences in pre-response theta between threatening and non-threatening conditions at the Preparation and Decision stages to be similar to that of the Contextual Prop and Action experiments, respectively. Generally, we found this to be the case.

At the Preparation stage, theta increase was greatest in the non-threatening condition over fronto-central electrodes for all groups (although not significant in the Younger Novice group), closely matching observations from the Contextual Prop experiments. We had predicted that the source of these differences would be the ACC. Comparison in source space showed that for both Matched Novice and AFO groups, theta power increase in the ACC was greatest in non-threatening conditions, supporting our hypothesis. Note, this finding was only borderline significant in the AFO group. In addition, greater theta power was observed in primary motor areas for both groups and in visual cortex for the Matched Novices only.

At the Decision stage, sensor level analysis revealed a consistent pattern of activity across all groups. There was a greater increase in left central and left temporo-parietal theta and delta in the Attack condition versus the Surrender condition. Simultaneously, we observed an increase in central theta that was greater in the Surrender condition than the Attack condition. This pattern of activity suggests a consistent dipolar source of activity, to which our sensor level analysis was not sensitive. Unfortunately, source analysis did not convincingly identify the source of this activity. A single cluster suggesting greater theta at the left pre-central gyrus was for the AFO group, but this did not fit the findings at sensor level.

We made two additional predictions related to the introduction of an expertise manipulation and graded level of threat in the current experiment. First, based on findings from studies on open skill athletes (Di Russo et al., 2006; You et al., 2018), we expected to find a greater increase in frontal-midline theta activity in the AFO group than the Matched Novice group when inhibiting response to threat. We did not find any differences between AFOs and Matched Novices in non-threatening scenarios. Rather, we found differences between these groups when they responded to threat. At the Preparation stage we found that AFOs had significantly greater increases in midline theta than Matched Novices when they equipped a firearm, but the centre of the cluster was more posterior than expected. Using the same comparison, we found that AFOs had significantly greater frontal-midline delta increases. At the Decision stage, no clusters were found when comparing AFO with Matched Novice theta and delta activity before discharging a firearm. Our second prediction was that AFOs would be better able to differentiate threat levels and this would be reflected by greater difference in EEG between Knife and Handgun conditions at the Preparation stage. We found that AFOs had a greater difference in delta activity than Novices between these conditions over right central and occipital electrodes.

Chapter 6: Discussion

6.1. Overview

Much of the work presented in this thesis was done to ensure that the results of our final experiment would be both internally and ecologically valid. In Chapter 3, I showed how we developed neuro-VR methods and used them to create police firearms training scenarios which could be treated as experimental trials. Next, in Chapter 4, I presented our work validating this approach by collecting data from novice participants completing two experiments based on our 'shoot/don't-shoot' paradigm. Here, we replicated behavioural results from other 'shoot/don't-shoot' experiments (Correll et al., 2002, 2006; Fleming et al., 2010) and measured differences in EEG between responses to threatening and non-threatening scenarios that fit with findings from comparable Go/No-Go tasks (Nieuwenhuis et al., 2003). However, the work presented in Chapters 3 and 4 did not directly address our main research question: what effect does expertise have on police decision making? In our final experiment, presented in Chapter 5, we addressed this question by collecting data from expert AFOs and novice control groups as they completed a series of scenarios. These scenarios were more complicated than any we had used previously and were developed as a combination of our two earlier experiments.

In this chapter, I will first discuss the findings from that experiment as they relate to our hypotheses, alongside findings from exploratory analysis. Following this, I will critically evaluate our methods and suggest areas for future research and development. Finally, I will look ahead to where research on police decision making should go next and, more generally, how neuro-VR can help others address research questions in their own fields.

6.2. Key findings

The key findings from our final experiment related to differences in pre-response EEG between expert AFOs and the control group. Here, we expected to replicate findings from investigations of open skill athletes that suggested experts would have greater activity in the ACC than novices when inhibiting a response (Di Russo et al., 2006; You et al., 2018). We took a different approach to these studies by analysing response-locked changes in oscillatory power, rather than stimulus-locked ERPs. This was necessary because of the continuous (as opposed to discrete) properties of our stimuli. Our findings suggested an effect of expertise on activity in the ACC, but not for conditions where they inhibited a response. Rather, we found differences in theta and delta activity between AFOs and the control group when they did respond. Specifically, responding by equipping a firearm in response to a threatening virtual human. While we had expected differences in theta to be over frontal-midline electrodes, their location was distributed and more central. Differences in delta activity were clearly identifiable as frontal-midline activity. Interestingly, the differences in topography between theta and delta we observed (theta more central; delta more frontal) closely matched those reported by Lockhart et al. (2019) as the difference between Go and No-Go conditions. In the context of Go/No-Go tasks, theta and delta activity have not been disambiguated and are likely the product of the same or similar activity in the brain (Harper et al., 2014; Huster et al., 2013; Lockhart et al., 2019; Schmiedt-Fehr et al., 2011). This is supported by our own data where theta and delta covary, as shown by similar directions of both average effects within conditions (see Figure 21), and differences between conditions in almost all comparisons. Changes in theta and delta activity may have varied in magnitude and significance, but generally they were similar. Therefore, we can assume that the greater theta and delta activity observed in AFOs reflect a single source of activity.

Had differences in theta and delta been found for the contrast between inhibitory responses (i.e. not equipping a firearm in response to a non-threatening virtual human), we would have taken this to mean that experts' increased performance was due to better response inhibition (Di Russo et al., 2006; You et al., 2018). We could speculate that when equipping one firearm (e.g. Glock) participants had to inhibit equipping the other (e.g. Taser) and that the difference between groups reflects that. However, this conclusion does not fit with our broader findings, as we did not identify any differences in the non-threatening condition, when we were more certain participants had to inhibit their response to equip a firearm. More likely, the differences we observed were not related to response inhibition.

Interestingly, the finding that AFOs have greater frontal-midline theta when responding to a threatening stimulus replicates findings from, to my knowledge, the only other 'shoot/don't-shoot' study to manipulate expertise and measure EEG (Johnson et al., 2014). They interpreted the difference as indicative of greater alertness in experts. Note, the methods used in their study meant that direct comparisons could not be made with our results, as discussed in Chapter 5. Nonetheless, we can interpret our findings in a similar manner. Theta and delta activity are associated with motivation and orientation, particularly towards threat (Balconi et al., 2009; Başar et al., 2001; DeLaRosa et al., 2014; Knyazev, 2012). Greater pre-response theta/delta in AFOs may reflect their greater attention when identifying and responding to threatening stimuli. Our findings, alongside those from Johnson et al. (2014), support this conclusion.

The second key finding from our final experiment was the greater differences in EEG in the expert AFO group versus the control group when identifying different levels of threat. While both groups generally had the same pattern of EEG when identifying and responding to either a knife or a handgun (increased frontal-midline theta/delta; decreased posterior alpha and central beta; see Figure 21), we were able to identify differences between these conditions within each group and then compare those differences between groups. Within the AFO group, increases in theta and delta activity were greatest when equipping a Glock in response to a handgun. While a similar trend could be seen within the control group, the difference was smaller. Between-subject comparison of these differences revealed that AFOs had greater differences in central delta activity. This finding further emphasises AFOs greater attention towards highly threatening stimuli and greater distinction between different threat levels.

6.3. Limitations and future directions of our research

Due to excellent support from the Tactical Training Centre we were able to collect a moderate sized sample of 27 AFO participants. By comparing their data with a novice control group, we were able to demonstrate that AFO expertise leads to improvements in performance, as explained by measurable differences in brain activity. The main limitation of our sample was that we treated all AFOs within a single expert group, regardless of individual differences in expertise. Therefore, our findings could not explicitly inform on policy related to selection and training, because these factors were not included in our experimental design. To do so would require collection of data from a stratified sample of individuals at different stages in their AFO training. If control of between-subject error was required, an alternative longitudinal study following a group of individuals throughout their training would also be desirable. Questions related to selection procedures and training could then be addressed directly. For example, what training is effective and how can it be improved? Can measures taken from neuroimaging benefit selection procedures? While outside of the scope of the research presented in this thesis, addressing these questions would be the rational next step in our research trajectory.

We made improvements between our early experiments and the final experiment with regards to the scenarios used. This was done by increasing the number of possible actions participants could take and the number of possible outcomes within scenarios. However, when compared to the plethora of different situations AFOs must resolve while carrying out their duties, these scenarios were still quite limited. As we were using a game engine to create and present our scenarios, it would have been relatively simple to increase the number of different possibilities within the scenarios. For instance, by introducing another prop, such as a baseball bat, or adding in more ambiguous animations. In doing so we could have made the task more realistic and engaging for participants. The reason we did not do this is because it would have complicated our experimental design and subsequent analysis. We discussed possible ways around this problem when deciding whether to use multiple virtual humans or not. Our conclusion was that using multiple virtual humans would only be sensible if all virtual humans were unique. Similarly, it would be possible to introduce many more props and animations to create threatening and non-threatening scenarios, but only if they were not repeated. Instead, each unique scenario would need to be categorised, just as the repeated scenarios we used were, by threat level (high, medium and no threat) and compliance (attacking or surrendering). In this way, the experiment could be made more realistic without any cost to experimental design or analysis. This may also have led to participants treating the virtual human as more real, and therefore eliciting more ecologically valid behaviour (Reader & Holmes, 2016). Their decision to shoot may then have depended on their reluctance to cause harm, in addition to their perception of threat. If we were to continue this research, it would be beneficial to make these improvements to scenarios.

After evaluating many neuro-VR options at the start of this research project, we decided to use a head-mounted display combined with EEG. Both the head-mounted display we used and EEG served our purposes well throughout the duration of the project. Both, however, had limitations. Limitations of the Oculus Rift CV1, and head-mounted displays more generally, were discussed in detail in Chapter 3. Briefly, the resolution and field of view of the Oculus Rift CV1 is limited which means the virtual reality it presents is less vivid than desirable. This was a necessary restriction of most head-mounted displays at the time so as to limit demand on graphical processing power and to allow for low latency updates of the display in response to user movement. While we believed that these limitations were acceptable when compared with alternatives, such as presenting stimuli on a screen, or using 360° videos, they were nonetheless limitations of our research. Fortunately, graphical processing and head-mounted display technology have advanced, and continue to advance, quickly since we began our research. Researchers starting a virtual reality project now (2020) have access to much more powerful virtual reality technology than we did when we started this project, circa 2017. Note that in some ways we benefited from the lower resolution of the Oculus Rift CV1. While generally beneficial, the use of more vivid virtual reality technology means that additional care is required when making stimuli for experiments. For example, if virtual humans are being presented their facial expressions will be visible. If the face is not animated, or animated poorly, then it may have a more negative effect on presence and co-presence (D. Roberts et al., 2006; Slater et al., 2000, p.) than a low-resolution face. In this sense, researchers much be careful of The Uncanny Valley (Mori [1970] in Mori et al. [2012]) as virtual reality technology improves.

We encountered several issues related to our use of EEG. Notably, we found that narrow-band high-frequency artefacts related to the head-mounted display and broadband muscle artefacts were prevalent across our data. These were not insurmountable, but they did limit our analysis to lower frequencies. Fortunately, for our research questions this was acceptable, as we were mostly interested in theta activity, well below the frequency range of these artefacts. However, in future our analysis may benefit from measurement of high-frequency signals, such as gamma. For example, our existing analysis could be supplemented by measurements of phase amplitude coupling between theta and gamma related to decision making (Amemiya & Redish, 2018). Alternatively, analysis of unfiltered, raw EEG, as in empirical mode decomposition (Burgess, 2012), may be required. To do so would require additional consideration of hardware, shielding and signal detection techniques. Regarding hardware, there are now many head-mounted displays available which can complement different types of analysis. For example, those using display technology with a high refresh rate (120Hz, [Pimax, Shanghai, China]) which is further outside frequencies of interest (Török et al., 2014). Additional shielding measures could include fibre optic display cables and direct current battery power supplies to isolate physical connections from line noise. The effectiveness of these methods has not yet been evaluated but should be considered in future research.

Much of our EEG analysis relied on the high sampling rate of the equipment and relatively high frequency of signals within EEG. Combined, these properties allowed us to measure changes in signal associated with activity in the brain with temporal resolution at the order of 10-100ms. This high temporal resolution was essential to distinguish brain activity associated with sequential stimuli and responses within a single trial and, crucially, between pre- and post-response activity (Puce & Hämäläinen, 2017). Any future work, or work by others using naturalistic stimuli, would benefit from using neuroimaging methods with similarly high temporal resolution. However, the spatial resolution of EEG is more limited, as we found when conducting our source analysis. Our results from source analysis were mostly congruent with results at sensor level and fitted with our predictions. Namely, we were able to localise frontal-midline theta activity to the ACC (Bekker et al., 2005; Nieuwenhuis et al., 2003). However, if our analysis was exploratory, with no prior expectation to find differences at the ACC, we would have considered the findings ambiguous, as the brain regions identified by cluster-based analysis were distributed beyond the ACC.

6.4. Future research on police decision making

While research into police decision making is not yet developed as a field, there does appear to be general agreement about how we can move it forward. Considering the differences between early ‘shoot/don’t-shoot’ paradigms (e.g. Correll et al. 2002) and more recent research (e.g. Johnson et al., 2018; Taylor, 2019), there has been a shift away from using static images to using video simulations as stimuli. Even more recently, researchers have used modern virtual reality technology in research related to shooting accuracy (Muñoz et al., 2020). This trend shows that researchers understand that vivid, interactive scenarios are necessary to ensure ecological validity when studying police officer behaviour (Hope, 2016). Of course, given the work presented in this thesis, we agree with this approach. Future research on police performance is unlikely to regress towards using simple stimuli. Rather, it should pursue advances in technology, just as training programmes in the police sector do (e.g. Adaptive VR Ltd. 2020).

Of the limited amount of research that has been conducted on police decision making, a large proportion has investigated the effects of bias. In particular, many studies have attempted to measure the effects of racism against black men in the USA on police officers' decision to shoot (Correll et al., 2002, 2006; James et al., 2016; Johnson et al., 2018; Plant & Peruche, 2005; Worrall et al., 2018). Throughout this thesis I have referred to some of these studies, but only where factors relevant to our own research, such as expertise or threat level, could be isolated from their other findings. Without this interpretation, very little research on police officer decision making would remain. While understanding the effect of bias on police decision making is important from a policy and societal point of view, we do not yet understand the basic processes of decision making that bias acts upon (Johnson et al., 2018). Future research into the effects of bias should first consider how these processes can be measured.

As we have found from our collaboration with the Tactical Training Centre, alongside reports from others (Suss & Boulton, 2019), police decision making is a complex and multifaceted process. We cannot yet realistically study all aspects of police decision making with quantitative methods. However, we can focus on what is currently achievable. There has been success from taking a reductionist approach and attempting to manipulate just one aspect of police decision making and measure its effect on a small number of outcomes. For example, prior information on shooting error (Johnson et al., 2018) and anxiety on distance perception (Nieuwenhuys, Cañal-Bruland, et al., 2012).

6.5. Future directions of neuro-VR

Having now completed the project, I still believe that our decision to use an Oculus Rift CV1 head-mounted display combined with EEG was the right one. Since that decision, both virtual reality and neuroimaging techniques have developed. Head-mounted displays have generally become more complex and now include more electronics. For example, many are now standalone devices, containing graphics processing hardware, or use wireless connections (e.g. Oculus Quest [Facebook, Inc., Menlo Park, CA, USA]). This presents new challenges for researchers who want to use the latest virtual reality equipment but must still minimise artefacts in their data and interface with other scientific equipment. To my knowledge, no manufacturers of head-mounted displays have developed products more suitable for neuroscience. Given the relatively small demand, it is unlikely they will in future. More likely, researchers will have to modify/develop head-mounted displays as applications demand them. For example, in other projects I have worked on, we have removed ferromagnetic materials from a head-mounted display and replaced active motion capture with passive alternatives (G. Roberts et al., 2019). On the 'neuro' side of neuro-VR, further advances in neuroimaging techniques, such as OP-MEG, fNIRS and mobile EEG, or other as yet unknown methods, will promote the field.

To accompany these advances in neuro-VR, improvements in signal processing to tackle inherently noisy datasets are needed. Were we to analyse our data again, new filtering techniques suitable for removing high frequency noise could be used. For example, Zapline (Cheveigné, 2019) could be adapted to model and remove high frequency noise sources, such as the unknown 52.1Hz artefact introduced by the head-mounted display. Alternatively, because electronic noise sources are so well defined, spectral interpolation could be used to reduce their effect (Leske & Dalal, 2019). This method works by reducing the power of a peak in PSD to the average of surrounding frequencies and then taking the inverse Fourier transform as a reconstruction of the data. One problem with this method is that it ignores macro-variation of the PSD, such as the 1/f-like trend. Alternatives methods of interpolation, such as modelling and subtracting peaks, may be more suitable, but have not yet been used for analysing neural timeseries data. Generally, it is essential that researchers using neuro-VR report details of their analysis and the problem solving involved so that others can learn from them (Klug & Gramann, 2020; Nenna et al., 2020; Wunderlich & Gramann, 2020). As neuro-VR becomes more widely used and accessible, researchers will be more motivated to tackle these problems and contribute to common analysis protocols.

The analysis techniques we applied to our EEG data were not specifically tailored to the naturalistic experiments we were running. Rather, we took the approach of carefully designing experiments so that our data would be amenable to more traditional analysis techniques. This was possible because the behaviour we were observing had clear, measurable events. However, there are many other behaviours that would be interesting to investigate using neuro-VR methods that cannot be reduced to the same extent. For these investigations, novel analysis techniques are required. For example, analysing off-task neural activity between events for related sequences may be useful when the precise timings of events is unknown (Liu et al., 2020). In other situations, identifying changes in neural activity related to events from the data itself may be another way to identify event-related epochs

(Geerligs et al., 2020). Alternatively, incorporation of additional measures, such as eye tracking and motion capture, allows for data driven extraction of events, such as blinks, saccades and changes in gait, to be analysed (Wunderlich & Gramann, 2020). These techniques, and others, may have been useful for analysing our data if we had more open-ended scenarios. We would have benefited from them if the environment or virtual humans responded more to participants' actions. For example, retreating or attacking in response to participants raising their firearm towards them.

6.6. Conclusion

In this thesis I have shown how, when combined, EEG and virtual reality can be used to produce and record behaviours in response to naturalistic stimuli while simultaneously recording high quality brain activity. We used this neuro-VR approach to investigate police officers' decisions regarding the use of stopping force in scenarios based on AFO training and experiences. In doing so, we achieved our two aims: to further develop, test and report on neuro-VR methods, and to investigate the effects of expertise on decision making. Crucially, the novelty of our approach allowed us to maintain high methodological standards, including the use of suitable control groups and control over independent variables, while using an ecologically valid experimental paradigm. Our key findings were that increased performance of AFOs, when compared with a novice control group, can be explained by differences in pre-response brain activity. Specifically, greater increases in frontal-midline theta and central delta activity before equipping a firearm. Further, we found that AFOs had greater differences in delta activity when responding to different levels of threat. These findings suggest that differences in performance between experts and novices may be due to greater attention towards threat. Further investigation of expert decision making should build on our use of naturalistic stimuli and expert participants to ensure that findings are ecologically valid. With increasing accessibility of modern game engines and virtual reality technology, this approach will be beneficial to researchers in many fields where lack of ecological validity has been problematic.

References

Abrash, M. (2012, December). Latency – the sine qua non of AR and VR. *Ramblings in Valve Time*. http://blogs.valvesoftware.com/abrash/latency-the-sine-qua-non-of-ar-and-vr/

Abtahi, P., Gonzalez-Franco, M., Ofek, E., & Steed, A. (2019). I'm a giant: Walking in large virtual environments at high speed gains. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. https://doi.org/10.1145/3290605.3300752

Adaptive VR Ltd. (2020). *AVRT – Adaptive Virtual Reality Training*. AVRT Adaptive Virtual Reality Training. https://avrt.training/

Akinola, M., & Mendes, W. B. (2012). Stress-induced cortisol facilitates threat-related decision making among police officers. *Behavioral Neuroscience*, *126*(1), 167–174. https://doi.org/10.1037/a0026657

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*(7), 266–271. https://doi.org/10.1016/0166-2236(90)90107-L

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, *6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

Amemiya, S., & Redish, A. D. (2018). Hippocampal Theta-Gamma Coupling Reflects State-Dependent Information Processing in Decision Making. *Cell Reports*, *22*(12), 3328–3338. https://doi.org/10.1016/j.celrep.2018.02.091

Angel, H., Adams, B. D., Brown, A., Flear, C., Mangan, B., Morten, A., & Ste-Croix, C. (2012). *Review of the Skills Perishability of Police "Use of Force" Skills*. HumanSystems Incorporated. http://www.policecouncil.ca/wp-content/uploads/2013/03/Police-skills-perishability-final-Feb-2012.pdf

Anthes, C., García-Hernández, R. J., Wiedemann, M., & Kranzlmüller, D. (2016). State of the art of virtual reality technology. *IEEE Aerospace Conference Proceedings*, *2016-June*. https://doi.org/10.1109/AERO.2016.7500674

Armstrong, J., Clare, J., & Plecas, D. (2014). Monitoring the impact of scenario-based use-of-force simulations on police heart rate: Evaluating the royal Canadian mounted police skills refresher program. *Western Criminology Review*, *15*(1), 51–59.

Asada, H., Fukuda, Y., Tsunoda, S., Yamaguchi, M., & Tonoike, M. (1999). Frontal midline theta rhythms reflect alternative activation of prefrontal cortex and anterior cingulate cortex in humans. *Neuroscience Letters*, *274*(1), 29–32. https://doi.org/10.1016/S0304-3940(99)00679-5

Ashburner, J., & Friston, K. J. (1999). Using Basis Functions. *Human Brain Mapping*, *266*, 254–266.

Aspell, J. E., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2013). Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception. *Psychological Science*, *24*(12), 2445–2453. https://doi.org/10.1177/0956797613498395

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience & Biobehavioral Reviews*, *44*, 76–93. https://doi.org/10.1016/j.neubiorev.2012.07.006

Bacchetti, P., Deeks, S. G., & McCune, J. M. (2011). Breaking Free of Sample Size Dogma to Perform Innovative Translational Research. *Science Translational Medicine*, *3*(87), 87ps24. https://doi.org/10.1126/scitranslmed.3001628

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*, *59*(4), 381–384. https://doi.org/10.1103/PhysRevLett.59.381

Balconi, M., Brambilla, E., & Falbo, L. (2009). Appetitive vs. Defensive responses to emotional cues. Autonomic measures and brain oscillation modulation. *Brain Research*, *1296*, 72–84. https://doi.org/10.1016/j.brainres.2009.08.056

Bandettini, P. A. (2009). What's new in neuroimaging methods? *Annals of the New York Academy of Sciences*, *1156*, 260–293. https://doi.org/10.1111/j.1749-6632.2009.04420.x

Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. *The Journal of Abnormal and Social Psychology*, *66*(1), 3–11. https://doi.org/10.1037/h0048687

Barry, D. N., Tierney, T. M., Holmes, N., Boto, E., Roberts, G., Leggett, J., Bowtell, R., Brookes, M. J., Barnes, G. R., & Maguire, E. A. (2019). Imaging the human hippocampus with optically-pumped magnetoencephalography. *NeuroImage*, *203*, 116192. https://doi.org/10.1016/j.neuroimage.2019.116192

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, *39*(2), 241–248. https://doi.org/10.1016/S0167-8760(00)00145-8

Baumgartner, T., Speck, D., Wettstein, D., Masnari, O., Beeli, G., & Jäncke, L. (2008). Feeling present in arousing virtual reality worlds: Prefrontal brain regions differentially orchestrate presence experience in adults and children. *Frontiers in Human Neuroscience*, *2*. https://doi.org/10.3389/neuro.09.008.2008

Baumgartner, T., Valko, L., Esslen, M., & Jäncke, L. (2006). Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study. *Cyberpsychology and Behavior*, *9*(1), 30–45. https://doi.org/10.1089/cpb.2006.9.30

Bayliss, J. D., & Ballard, D. H. (2000). A virtual reality testbed for brain-computer interface research. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 188–190. https://doi.org/10.1109/86.847811

Beach, L. R., & Mitchell, T. R. (1978). A Contingency Model for the Selection of Decision Strategies. *Academy of Management Review*, *3*(3), 439–449. https://doi.org/10.5465/amr.1978.4305717

Bebko, A. O., & Troje, N. (2020). BmlTUX: Design and control of experiments in virtual reality and beyond. *PsyArXiv*. https://doi.org/10.31234/osf.io/arvkf

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience*, *19*(13), 5473–5481. https://doi.org/10.1523/jneurosci.19-13-05473.1999

Behr, K. M., Nosper, A., Klimmt, C., & Hartmann, T. (2005). Some practical considerations of ethical issues in VR research. *Presence: Teleoperators and Virtual Environments*, *14*(6), 668–676. https://doi.org/10.1162/105474605775196535

Bekker, E. M., Kenemans, J. L., & Verbaten, M. N. (2005). Source analysis of the N2 in a cued Go/NoGo task. *Cognitive Brain Research*, *22*(2), 221–231. https://doi.org/10.1016/j.cogbrainres.2004.08.011

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Benikos, N., Johnstone, S. J., & Roodenrys, S. J. (2013). Varying task difficulty in the Go/Nogo task: The effects of inhibitory control, arousal, and perceived effort on ERP components. *International Journal of Psychophysiology*, *87*(3), 262–272. https://doi.org/10.1016/j.ijpsycho.2012.08.005

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Berger, C. C., Gonzalez-Franco, M., Ofek, E., & Hinckley, K. (2018). The uncanny valley of haptics. *Science Robotics*, *3*(17). https://doi.org/10.1126/scirobotics.aar7010

Berkman, M. I., & Akan, E. (2019). Presence and Immersion in Virtual Reality. In N. Lee (Ed.), *Encyclopedia of Computer Graphics and Games* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-08234-9_162-1

Biggs, A. T., Cain, M. S., & Mitroff, S. R. (2015). Cognitive Training Can Reduce Civilian Casualties in a Simulated Shooting Environment. *Psychological Science*, *26*(8), 1164–1176. https://doi.org/10.1177/0956797615579274

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication*, *3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

Blacker, K. J., Pettijohn, K. A., Roush, G., & Biggs, A. T. (2020). Measuring Lethal Force Performance in the Lab: The Effects of Simulator Realism and Participant Experience. *Human Factors*, 0018720820916975. https://doi.org/10.1177/0018720820916975

Blaha, C. D., Allen, L. F., Das, S., Inglis, W. L., Latimer, M. P., Vincent, S. R., & Winn, P. (1996). Modulation of dopamine efflux in the nucleus accumbens after cholinergic stimulation of the ventral tegmental area in intact, pedunculopontine tegmental nucleus-lesioned, and laterodorsal tegmental nucleus-lesioned rats. *Journal of Neuroscience*, *16*(2), 714–722. https://doi.org/10.1523/JNEUROSCI.16-02-00714.1996

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., Muñoz, L. D., Mullinger, K. J., Tierney, T. M., Bestmann, S., Barnes, G. R., Bowtell, R., & Brookes, M. J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature*, *555*(7698), 657–661. https://doi.org/10.1038/nature26147

Botvinick, M. M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. https://doi.org/10.1038/35784

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Boulic, R., Capin, T., Huang, Z., Kalra, P., Lintermann, B., Magnenat-Thalmann, N., Moccozet, L., Molet, T., Pandzic, I., Saar, K., Schmitt, A., Shen, J., & Thalmann, D. (1995). The HUMANOID Environment for Interactive Animation of Multiple Deformable Human Characters. *Computer Graphics Forum*, *14*(3), 337–348. https://doi.org/10.1111/j.1467-8659.1995.cgf143_0337.x

Boulton, L., & Cole, J. (2016). Adaptive flexibility: Examining the role of expertise in the decision making of authorized firearms officers during armed confrontation. *Journal of Cognitive Engineering and Decision Making*, *10*(3), 291–308. https://doi.org/10.1177/1555343416646684

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: Neural markers of low alertness during mind wandering. *NeuroImage*, *54*(4), 3040–3047. https://doi.org/10.1016/j.neuroimage.2010.10.008

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J., & Markowitsch, H. J. (2005). Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: Associations with executive functions. *Neuropsychology*, *19*(3), 267–277. https://doi.org/10.1037/0894-4105.19.3.267

Brand, M., Grabenhorst, F., Starcke, K., Vandekerckhove, M. M. P., & Markowitsch, H. J. (2007). Role of the amygdala in decisions under ambiguity and decisions under risk: Evidence from patients with Urbach-Wiethe disease. *Neuropsychologia*, *45*(6), 1305–1317. https://doi.org/10.1016/j.neuropsychologia.2006.09.021

Brevers, D., Dubuisson, E., Dejonghe, F., Dutrieux, J., Petieau, M., Cheron, G., Verbanck, P., & Foucart, J. (2018). Proactive and Reactive Motor Inhibition in Top Athletes Versus Nonathletes. *Perceptual and Motor Skills*, *125*. https://doi.org/10.1177/0031512517751751

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. (2020). *The timing mega-study: Comparing a range of experiment generators, both lab-based and online*. https://doi.org/10.31234/osf.io/d6nu5

Brigadoi, S., Ganglani, A., Zhao, H., Zhao, H., Cooper, R. J., & Cooper, R. J. (2019). Integrating motion sensing and wearable, modular high-density diffuse optical tomography: Preliminary results. *Clinical and Preclinical Optical Diagnostics II (2019), Paper 11074_4*, 11074_4. https://doi.org/10.1117/12.2527197

Brisinda, D., Venuti, A., Cataldi, C., Efremov, K., Intorno, E., & Fenici, R. (2015). Real-time Imaging of Stress-induced Cardiac Autonomic Adaptation During Realistic Force-on-force Police Scenarios. *Journal of Police and Criminal Psychology*, *30*(2), 71–86. https://doi.org/10.1007/s11896-014-9142-5

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2019). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01242-0

Bruna, R., Zani, A., Taddei, F., & Pesce, C. (1992). Chronometric aspects of information processing in high level fencers as compared to non-athletes: An ERPs and RT study. *Journal of Human Movement Studies*, *23*, 17–28.

Buckland, S. T. (1984). Monte Carlo Confidence Intervals. *Biometrics*, *40*(3), 811–817. JSTOR. https://doi.org/10.2307/2530926

Burgess, A. P. (2012). Towards a Unified Understanding of Event-Related Changes in the EEG: The Firefly Model of Synchronization through Cross-Frequency Phase Modulation. *PLoS ONE*, *7*(9), 1–21. https://doi.org/10.1371/journal.pone.0045630

Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A.-S. T. (2005). Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research*, *5*(4), 475–497. https://doi.org/10.1177/1468794105056924

Cabeza, R., Prince, S. E., Daselaar, S. M., Greenberg, D. L., Budde, M., Dolcos, F., LaBar, K. S., & Rubin, D. C. (2004). Brain Activity during Episodic Retrieval of Autobiographical and Laboratory Events: An fMRI Study using a Novel Photo Paradigm. *Journal of Cognitive Neuroscience*, *16*(9), 1583–1594. https://doi.org/10.1162/0898929042568578

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Callejas, A., Lupiáñez, J., & Tudela, P. (2004). The three attentional networks: On their independence and interactions. *Brain and Cognition*, *54*(3), 225–227. https://doi.org/10.1016/j.bandc.2004.02.012

Caloyannides, M. A. (1974). Microcycle spectral estimates of 1/f noise in semiconductors. *Journal of Applied Physics*, *45*(1), 307–316. https://doi.org/10.1063/1.1662977

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370. https://doi.org/10.1007/BF00122575

Carey, M., Crucianelli, L., Preston, C., & Fotopoulou, A. (2019). The Effect of Visual Capture Towards Subjective Embodiment Within the Full Body Illusion. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-39168-4

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., Asmundson, G. J. G., Carlbring, P., & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders*, *61*(June 2018), 27–36. https://doi.org/10.1016/j.janxdis.2018.08.003

Carr, G. D., & White, N. M. (1983). Conditioned place preference from intra-accumbens but not intra-caudate amphetamine injections. *Life Sciences*, *33*(25), 2551–2557. https://doi.org/10.1016/0024-3205(83)90165-0

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Charles, M. T., & Copay, A. G. (2003). Acquisition of Marksmanship and Gun Handling Skills through Basic Law Enforcement Training in an American Police Department. *International Journal of Police Science & Management*, *5*(1), 16–30. https://doi.org/10.1350/ijps.5.1.16.11245

Chénéchal, M. L., & Chatel-Goldman, J. (2018). *HTC Vive Pro Time Performance Benchmark for Scientific Research*. 81–84. https://doi.org/10.2312/egve.20181318

Cherni, H., Métayer, N., & Souliman, N. (2020). Literature review of locomotion techniques in virtual reality. *International Journal of Virtual Reality*, *20*(1), 1–20. https://doi.org/10.20870/IJVR.2020.20.1.3183

Cheveigné, A. de. (2019). ZapLine: A simple and effective method to remove power line artifacts. *BioRxiv*, 782029. https://doi.org/10.1101/782029

Christensen-Szalanski, J. J. J. (1978). Problem solving strategies: A selection mechanism, some implications, and some data. *Organizational Behavior and Human Performance*, *22*(2), 307–323. https://doi.org/10.1016/0030-5073(78)90019-3

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, *9*(NOV), 1–20. https://doi.org/10.3389/fpsyg.2018.02086

Cockbain, E., & Knutsson, J. (2015). *Applied Police Research: Challenges and opportunities*. Routledge, Abingdon, UK.

College of Policing. (2013). *Deployment of AFOs*. https://www.app.college.police.uk/app-content/armed-policing/use-of-force-firearms-and-less-lethal-weapons/deployment-of-afos/

College of Policing. (2014). *Code of Ethics: A Code of Practice for the Principles and Standards of Professional Behaviour for the Policing Profession of England and Wales*.

College of Policing. (2020). *Code of Practice on Armed Policing and Police use of Less Lethal Weapons*.

Collins, D. L., Zijdenbos, A. P., Baaré, W. F. C., & Evans, A. C. (1999). ANIMAL+INSECT: Improved Cortical Structure Segmentation. In A. Kuba, M. Šáamal, & A. Todd-Pokropek (Eds.), *Information Processing in Medical Imaging* (pp. 210–223). Springer. https://doi.org/10.1007/3-540-48714-X_16

Conn, C., Lanier, J., Minsky, M., Fisher, S., & Druin, A. (1989). Virtual environments and interactivity: Windows to the future. *ACM SIGGRAPH Computer Graphics*, *23*(5), 7–18. https://doi.org/10.1145/77277.77278

Connelly, M., Suss, J., & DiBello, L. (2019). Improving Expertise in Local Law Enforcement: Utilizing Virtual Environments to Assess Officer Performance and Standardize Training

Procedures. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2144–2148. https://doi.org/10.1177/1071181319631387

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer’s dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Correll, J., Urland, G. R., & Ito, T. A. (2006). Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*(1), 120–128. https://doi.org/10.1016/j.jesp.2005.02.006

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM*, *35*(6), 64–72. https://doi.org/10.1145/129888.129892

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable Neural Components in the Processing of Black and White Faces. *Psychological Science*, *15*(12), 806–813. https://doi.org/10.1111/j.0956-7976.2004.00760.x

Currey, J., & Botti, M. (2003). Naturalistic Decision Making: A Model to Overcome Methodological Challenges in the Study of Critical Care Nurses’ Decision Making About Patients’ Hemodynamic Status. *American Journal of Critical Care*, *12*(3), 206–211. https://doi.org/10.4037/ajcc2003.12.3.206

Dalal, S. S., Rampp, S., Willomitzer, F., & Ettl, S. (2014). Consequences of EEG electrode position error on ultimate beamformer source reconstruction performance. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00042

Damasio, A. R., Everitt, B. J., Bishop, D., Roberts, A. C., Robbins, T. W., & Weiskrantz, L. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28–43. https://doi.org/10.1016/j.jml.2018.07.005

de Gelder, B., Kätsyri, J., & de Borst, A. W. (2018). Virtual reality and the new psychophysics. *British Journal of Psychology*, *109*(3), 421–426. https://doi.org/10.1111/bjop.12308 de la Rosa, S., & Breidt, M. (2018). Virtual reality: A new track in psychological research. *British Journal of Psychology*, *109*(3), 427–430. https://doi.org/10.1111/bjop.12302

Dedovic, K., D'Aguiar, C., & Pruessner, J. C. (2009). What Stress Does to Your Brain: A Review of Neuroimaging Studies. *The Canadian Journal of Psychiatry*, *54*(1), 6–15. https://doi.org/10.1177/070674370905400104

DeFanti, T. A., Acevedo, D., Ainsworth, R. A., Brown, M. D., Cutchin, S., Dawe, G., Doerr, K.-U., Johnson, A., Knox, C., Kooima, R., Kuester, F., Leigh, J., Long, L., Otto, P., Petrovic, V., Ponto, K., Prudhomme, A., Rao, R., Renambot, L., … Wickham, G. (2011). The future of the CAVE. *Central European Journal of Engineering*, *1*(1), 16–37. https://doi.org/10.2478/s13531-010-0002-5

DeLaRosa, B. L., Spence, J. S., Shakal, S. K. M., Motes, M. A., Calley, C. S., Calley, V. I., Hart, J., & Kraut, M. A. (2014). Electrophysiological spatiotemporal dynamics during implicit visual threat processing. *Brain and Cognition*, *91*, 54–61. https://doi.org/10.1016/j.bandc.2014.08.003

Dempsey, M. F., Condon, B., & Hadley, D. M. (2002). MRI safety review. *Seminars in Ultrasound, CT and MRI*, *23*(5), 392–401. https://doi.org/10.1016/S0887-2171(02)90010-7

Dexter, J. P., Prabakaran, S., & Gunawardena, J. (2019). A Complex Hierarchy of Avoidance Behaviors in a Single-Cell Eukaryote. *Current Biology*, *0*(0). https://doi.org/10.1016/j.cub.2019.10.059

Di Russo, F., Taddei, F., Apnile, T., & Spinelli, D. (2006). Neural correlates of fast stimulus discrimination and response selection in top-level fencers. *Neuroscience Letters*, *408*(2), 113–118. https://doi.org/10.1016/j.neulet.2006.08.085

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How Does the Brain Solve Visual Object Recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, *6*(JAN), 1–9. https://doi.org/10.3389/fpsyg.2015.00026

Draganski, B., Kherif, F., Klöppel, S., Cook, P. A., Alexander, D. C., Parker, G. J. M., Deichmann, R., Ashburner, J., & Frackowiak, R. S. J. (2008). Evidence for Segregated and Integrative Connectivity Patterns in the Human Basal Ganglia. *Journal of Neuroscience*, *28*(28), 7143–7152. https://doi.org/10.1523/JNEUROSCI.1486-08.2008

Ejelöv, E., & Luke, T. J. (2020). “Rarely safe to assume”: Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937. https://doi.org/10.1016/j.jesp.2019.103937

Ekman, E. (2015). Development, delivery, and evaluation of a pilot stress reduction, emotion regulation, and mindfulness training for juvenile justice officers. *Journal of Juvenile Justice*, *4*(2), 71–94.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75. https://doi.org/10.1007/BF01115465

Elliott, Gemma., Isaac, C. L., & Muhlert, N. (2014). Measuring forgetting: A critical review of accelerated long-term forgetting studies. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *54*(100), 16–32. https://doi.org/10.1016/j.cortex.2014.02.001

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, *75*(4), 643. https://doi.org/10.2307/1884324

Enman, N. M., Arthur, K., Ward, S. J., Perrine, S. A., & Unterwald, E. M. (2015). Anhedonia, Reduced Cocaine Reward, and Dopamine Dysfunction in a Rat Model of Posttraumatic Stress Disorder. *Biological Psychiatry*, *78*(12), 871–879. https://doi.org/10.1016/j.biopsych.2015.04.024

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. *Journal of Personality and Social Psychology*, *71*(2), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

Ermer, J. J. (2001). *Realistic forward modeling and source complexity reduction methods for EEG and MEG* [Dissertation, University of Southern California]. http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll16/id/96637

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statistical Science*, *19*(4), 676–685. https://doi.org/10.1214/088342304000000396

Ettl, S., Rampp, S., Fouladi-Movahed, S., Dalal, S. S., Willomitzer, F., Arold, O., Stefan, H., & Häusler, G. (2013). Improved EEG source localization employing 3D sensing by 'Flying Triangulation'. *Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, *8791*, 87910V. https://doi.org/10.1117/12.2020533

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Farmer, H., Tajadura-Jiménez, A., & Tsakiris, M. (2012). Beyond the colour of my skin: How skin colour affects the sense of body-ownership. *Consciousness and Cognition*, *21*(3), 1242–1256. https://doi.org/10.1016/j.concog.2012.04.011

Feldstein, I. T. (2019). Impending Collision Judgment from an Egocentric Perspective in Real and Virtual Environments: A Review. *Perception*, *48*(9), 769–795. https://doi.org/10.1177/0301006619861892

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*(12), 1643–1646. https://doi.org/10.1038/nn1608

Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology*, *3*(10). https://doi.org/10.1371/journal.pbio.0030342

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, *112*(3), 536–544. https://doi.org/10.1016/S1388-2457(00)00533-2

Fishburn, P. C. (1982). Nontransitive measurable utility. *Journal of Mathematical Psychology*, *26*(1), 31–67. https://doi.org/10.1016/0022-2496(82)90034-7

Fishburn, P. C. (1989). Non-transitive measurable utility for decision under uncertainty. *Journal of Mathematical Economics*, *18*(2), 187–207. https://doi.org/10.1016/0304-4068(89)90021-9

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327–358. https://doi.org/10.1037/h0061470

Fleming, K. K., Bandy, C. L., & Kimble, M. O. (2010). Decisions to shoot in a weapon identification task: The influence of cultural stereotypes and perceived threat on false positive errors. *Social Neuroscience*, *5*(2), 201–220. https://doi.org/10.1080/17470910903268931

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *Supplement 1*(47), S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Fontaine, G. (1992). The Experience of a Sense of Presence in Intercultural and International Encounters. *Presence: Teleoperators and Virtual Environments*, *1*(4), 482–490. https://doi.org/10.1162/pres.1992.1.4.482

Fredericksen, R. E., Bex, P. J., & Verstraten, F. A. J. (1997). How big is a Gabor patch, and why should we care? *JOSA A*, *14*(1), 1–12. https://doi.org/10.1364/JOSAA.14.000001

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, *5*(8), 625–632. https://doi.org/10.1016/S2215-0366(18)30226-8

Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & IJsselsteijn, W. (2000). Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence: Teleoperators and Virtual Environments*, *9*(2), 149–164. https://doi.org/10.1162/105474600566691

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. https://doi.org/10.1002/hbm.460020402

Fulvio, J. M., & Rokers, B. (2017). Use of cues in virtual reality depends on visual feedback. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-16161-3

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting Eyewitness Evidence: Examining the Efficacy of a Self-Administered Interview Tool. *Law and Human Behavior*, *33*(4), 298–307. https://doi.org/10.1007/s10979-008-9146-8

Gabriel, S., Lau, R. W., & Gabriel, C. (1996). The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. *Physics in Medicine and Biology*, *41*(11), 2251–2269. https://doi.org/10.1088/0031-9155/41/11/002

Gajewski, P. D., & Falkenstein, M. (2013). Effects of task complexity on ERP components in Go/Nogo tasks. *International Journal of Psychophysiology*, *87*(3), 273–278. https://doi.org/10.1016/j.ijpsycho.2012.08.007

Gee, F. C., Browne, W. N., & Kawamura, K. (2005). Uncanny valley revisited. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 151–157. https://doi.org/10.1109/ROMAN.2005.1513772

Geerligs, L., Gerven, M. van, & Güçlü, U. (2020). Detecting neural state transitions underlying event segmentation. *BioRxiv*, 2020.04.30.069989. https://doi.org/10.1101/2020.04.30.069989

George, C., Janssen, P., Heuss, D., & Alt, F. (2019). Should I Interrupt or Not? Understanding Interruptions in Head-Mounted Display Settings. *Proceedings of the 2019 on Designing Interactive Systems Conference*, 497–510. https://doi.org/10.1145/3322276.3322363

Gibson, J. J. (2014). The Ecological Approach to Visual Perception. In *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA. https://doi.org/10.4324/9781315740218

Gibson, W. (1984). *Neuromancer*. Gollancz, London, UK.

Giessing, L., Frenkel, M. O., Zinner, C., Rummel, J., Nieuwenhuys, A., Kasperk, C., Brune, M., Engel, F. A., & Plessner, H. (2019). Effects of Coping-Related Traits and Psychophysiological Stress Responses on Police Recruits' Shooting Behavior in Reality-Based Scenarios. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01523

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In *Simple heuristics that make us smart* (pp. 75–95). Oxford University Press, Oxford, UK.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, *6*(1), 100–121.

Gourlay, M. J., & Held, R. T. (2017). Head-Mounted-Display Tracking for Augmented and Virtual Reality. *Information Display*, *33*(1), 6–10. https://doi.org/10.1002/j.2637-496X.2017.tb00962.x

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology*, *91*(1), 22–29. https://doi.org/10.1016/j.ijpsycho.2013.09.003

Graybiel, A. M., & Grafton, S. T. (2015). The Striatum: Where Skills and Habits Meet. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021691. https://doi.org/10.1101/cshperspect.a021691

Grimshaw, M. (2009, January 1). The audio Uncanny Valley: Sound, fear and the horror game. *Games Computing and Creative Technologies: Conference Papers (Peer-Reviewed)*.

Gutshall, C. L., Hampton, D. P., Sebetan, I. M., Stein, P. C., & Broxtermann, T. J. (2017). The effects of occupational stress on cognitive performance in police officers. *Police Practice and Research*, *18*(5), 463–477. https://doi.org/10.1080/15614263.2017.1288120

Haber, S. N. (2017). Chapter 1—Anatomy and Connectivity of the Reward Circuit. In J.-C. Dreher & L. Tremblay (Eds.), *Decision Neuroscience* (pp. 3–19). Academic Press, San Diego, CA, USA. https://doi.org/10.1016/B978-0-12-805308-9.00001-4

Haber, S. N., & Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging. *Neuropsychopharmacology*, *35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Haber, S. N., Kunishio, K., Mizobuchi, M., & Lynd-Balta, E. (1995). The orbital and medial prefrontal circuit through the primate basal ganglia. *Journal of Neuroscience*, *15*(7), 4851–4867. https://doi.org/10.1523/JNEUROSCI.15-07-04851.1995

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage*, *20*, S18–S29. https://doi.org/10.1016/j.neuroimage.2003.09.013

Hajihosseini, A., & Holroyd, C. B. (2013). Frontal midline theta and N200 amplitude reflect complementary information about expectancy and outcome evaluation. *Psychophysiology*, *50*(6), 550–562. https://doi.org/10.1111/psyp.12040

Halder, T., Talwar, S., Jaiswal, A. K., & Banerjee, A. (2019). Quantitative evaluation in estimating sources underlying brain oscillations using current source density methods and beamformer approaches. *ENeuro*, *6*(4), 1–14. https://doi.org/10.1523/ENEURO.0170-19.2019

Hamilton, J. A., Lambert, G., Suss, J., & Biggs, A. T. (2019). Can Cognitive Training Improve Shoot/Don't-Shoot Performance? Evidence from Live Fire Exercises. *The American Journal of Psychology*, *132*(2), 179–194. JSTOR.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct Comparison of the Efficacy of Intuitive and Analytical Cognition in Expert Judgment. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-17*(5), 753–770. https://doi.org/10.1109/tsmc.1987.6499282

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of Social Interaction in Human Brain Function. *Neuron*, *88*(1), 181–193. https://doi.org/10.1016/j.neuron.2015.09.022

Hari, R., & Puce, A. (2017). *MEG-EEG Primer*. Oxford University Press, Oxford, UK.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, *7*(DEC), 1–10. https://doi.org/10.3389/fnint.2013.00083

Harper, J., Malone, S. M., & Bernat, E. M. (2014). Theta and delta band activity explain N2 and P3 ERP component activity in a go/no-go task. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *125*(1), 124–132. https://doi.org/10.1016/j.clinph.2013.06.025

Harris, K. R., Eccles, D. W., Freeman, C., & Ward, P. (2017). 'Gun! Gun! Gun!': An exploration of law enforcement officers' decision-making and coping under stress during actual events. *Ergonomics*, *60*(8), 1112–1122. https://doi.org/10.1080/00140139.2016.1260165

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, *57*(2), 243–259. JSTOR. https://doi.org/10.2307/1416950

Heimer, L., Switzer, R. D., & Van Hoesen, G. W. (1982). Ventral striatum and ventral pallidum: Components of the motor system? *Trends in Neurosciences*, *5*, 83–87. https://doi.org/10.1016/0166-2236(82)90037-6

Henkel, L. A. (2014). Point-and-Shoot Memories: The Influence of Taking Photos on Memory for a Museum Tour. *Psychological Science*, *25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Herrmann, M. J., Römmler, J., Ehlis, A.-C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, *20*(2), 294–299. https://doi.org/10.1016/j.cogbrainres.2004.02.013

Hertweck, S., Weber, D., Alwanni, H., Unruh, F., Fischbach, M., Latoschik, M. E., & Ball, T. (2019). Brain activity in virtual reality: Assessing signal quality of high-resolution EEG while using head-mounted displays. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 970–971. https://doi.org/10.1109/VR.2019.8798369

Heydrich, L., Aspell, J. E., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2018). Cardio-visual full body illusion alters bodily self-consciousness and tactile processing in somatosensory cortex. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-27698-2

Hill, R. M., Boto, E., Rea, M., Holmes, N., Leggett, J., Coles, L. A., Papastavrou, M., Everton, S. K., Hunt, B. A. E., Sims, D., Osborne, J., Shah, V., Bowtell, R., & Brookes, M. J. (2020). Multi-channel whole-head OPM-MEG: Helmet design and a comparison with a conventional system. *NeuroImage*, *219*, 116995. https://doi.org/10.1016/j.neuroimage.2020.116995

Hine, K. A., Porter, L. E., Westera, N. J., Alpert, G. P., & Allen, A. (2018). Exploring Police Use of Force Decision-Making Processes and Impairments Using a Naturalistic Decision-Making Approach. *Criminal Justice and Behavior*, *45*(11), 1782–1801. https://doi.org/10.1177/0093854818789726

Hollon, N. G., Burgeno, L. M., & Phillips, P. E. M. (2015). Stress effects on the neural substrates of motivated behavior. *Nature Neuroscience*, *18*(10), 1405–1412. https://doi.org/10.1038/nn.4114

Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow & Metabolism*, *16*(1), 7–22. https://doi.org/10.1097/00004647-199601000-00002

Holroyd, C. B. (2004). A note on the oddball N200 and the feedback ERN. *Neurophysiology*, *78*(1), 447–455.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Home Office. (2010). *Assessment of women in the police service*. (p. 28p., tabs.). Home Office.

Home Office. (2019a). *Police use of firearms statistics, England and Wales: April 2018 to March 2019*. 17.

Home Office. (2019b). *Police workforce, England and Wales, 31 March 2019 second edition*.

Homölle, S., & Oostenveld, R. (2019). Using a structured-light 3D scanner to improve EEG source modeling with more accurate electrode positions. *Journal of Neuroscience Methods*, *326*(February), 108378. https://doi.org/10.1016/j.jneumeth.2019.108378

Hope, L. (2016). Evaluating the Effects of Stress and Fatigue on Police Officer Response and Recall: A Challenge for Research, Training, Practice and Policy. *Journal of Applied Research in Memory and Cognition*, *5*(3), 239–245. https://doi.org/10.1016/j.jarmac.2016.07.008

Hope, L., Gabbert, F., & Fraser, J. (2013). Postincident conferring by law enforcement officers: Determining the impact of team discussions on statement content, accuracy, and officer beliefs. *Law and Human Behavior*, *37*(2), 117–127. https://doi.org/10.1037/lhb0000019

Hough, M., Jackson, J., Bradford, B., Myhill, A., & Quinton, P. (2010). Procedural Justice, Trust, and Institutional Legitimacy. *Policing: A Journal of Policy and Practice*, *4*(3), 203–210. https://doi.org/10.1093/police/paq027

Hoyet, L., Argelaguet, F., Nicole, C., & Lécuyer, A. (2016). "Wow! I Have Six Fingers!": Would You Accept Structural Changes of Your Hand in VR? *Frontiers in Robotics and AI*, *3*. https://doi.org/10.3389/frobt.2016.00027

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., & Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. *International Journal of Psychophysiology*, *87*(3), 217–233. https://doi.org/10.1016/j.ijpsycho.2012.08.001

iFIXIT. (2016, March 30). *Oculus Rift CV1 Teardown*. IFixit. https://www.ifixit.com/Teardown/Oculus+Rift+CV1+Teardown/60612

Inquest. (2020). *Fatal police shootings*. Inquest. https://www.inquest.org.uk/fatal-police-shootings

Ironside, M., Kumar, P., Kang, M.-S., & Pizzagalli, D. A. (2018). Brain mechanisms mediating effects of stress on reward sensitivity. *Current Opinion in Behavioral Sciences*, *22*, 106–113. https://doi.org/10.1016/j.cobeha.2018.01.016

James, L., James, S. M., & Vila, B. J. (2016). The Reverse Racism Effect. *Criminology & Public Policy*, *15*(2), 457–479. https://doi.org/10.1111/1745-9133.12187

Jasper, H. H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*(2), 371–375. https://doi.org/10.1016/0013-4694(58)90053-1

Jatoi, M. A., Kamel, N., Malik, A. S., & Faye, I. (2014). EEG based brain source localization comparison of sLORETA and eLORETA. *Australasian Physical and Engineering Sciences in Medicine*, *37*(4), 713–721. https://doi.org/10.1007/s13246-014-0308-3

Jin, X., & Costa, R. M. (2015). Shaping action sequences in basal ganglia circuits. *Current Opinion in Neurobiology*, *33*, 188–196. https://doi.org/10.1016/j.conb.2015.06.011

Joel, D., & Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: Open interconnected rather than closed segregated. *Neuroscience*, *63*(2), 363–379. https://doi.org/10.1016/0306-4522(94)90536-3

Johnson, D. J., Cesario, J., & Pleskac, T. J. (2018). How prior information and police experience impact decisions to shoot. *Journal of Personality and Social Psychology*, *115*(4), 601–623. https://doi.org/10.1037/pspa0000130

Johnson, J. B. (1925). The Schottky Effect in Low Frequency Circuits. *Physical Review*, *26*(1), 71–85. https://doi.org/10.1103/PhysRev.26.71

Johnson, J. G., & Raab, M. (2003). Take The First: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes*, *91*(2), 215–229. https://doi.org/10.1016/S0749-5978(03)00027-X

Johnson, R. R., Stone, B. T., Miranda, C. M., Vila, B., James, L., James, S. M., Rubio, R. F., & Berka, C. (2014). Identifying psychophysiological indices of expert vs. Novice performance in deadly force judgment and decision making. *Frontiers in Human Neuroscience*, *8*(JULY), 1–13. https://doi.org/10.3389/fnhum.2014.00512

Joseph, M. H., Datla, K., & Young, A. M. J. (2003). The interpretation of the measurement of nucleus accumbens dopamine by in vivo dialysis: The kick, the craving or the cognition? *Neuroscience and Biobehavioral Reviews*, *27*(6), 527–541. https://doi.org/10.1016/j.neubiorev.2003.09.001

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2019). Chapter 10—MoBI—Mobile Brain/Body Imaging. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics* (pp. 59–63). Academic Press. https://doi.org/10.1016/B978-0-12-811926-6.00010-5

Kaiser, J., Simon, N. A., Sauseng, P., & Schütz-Bosbach, S. (2019). Midfrontal neural dynamics distinguish between general control and inhibition-specific processes in the stopping of motor actions. *Scientific Reports*, *9*(1), 13054. https://doi.org/10.1038/s41598-019-49476-4

Karni, E. (2014). Axiomatic foundations of expected utility and subjective probability. In *Handbook of the Economics of Risk and Uncertainty* (1st ed., Vol. 1). Elsevier B.V., Oxford, UK. https://doi.org/10.1016/B978-0-444-53685-3.00001-5

Kayser, C., Einhäuser, W., & König, P. (2003). Temporal correlations of orientations in natural scenes. *Neurocomputing*, *52–54*, 117–123. https://doi.org/10.1016/S0925-2312(02)00789-0

Keenaghan, S., Bowles, L., Crawfurd, G., Thurlbeck, S., Kentridge, R. W., & Cowie, D. (2020). My body until proven otherwise: Exploring the time course of the full body illusion. *Consciousness and Cognition*, *78*, 102882. https://doi.org/10.1016/j.concog.2020.102882

Khalighinejad, N., Priestley, L., Jbabdi, S., & Rushworth, M. F. S. (2020). Human decisions about when to act originate within a basal forebrain–nigral circuit. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1921211117

Kisker, J., Gruber, T., & Schöne, B. (2019). Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: A study on human memory. *Current Psychology*. https://doi.org/10.1007/s12144-019-00257-2

Klatzky, R. L., Lederman, S. J., & Metzger, V. A. (1985). Identifying objects by touch: An “expert system”. *Perception & Psychophysics*, *37*(4), 299–302. https://doi.org/10.3758/BF03211351

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In *Decision making in action: Models and methods* (pp. 138–147). Ablex Publishing.

Klein, G. A. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Klein, G. A. (2015a). *Gary Klein—Affiliations*. Garyklein. https://www.gary-klein.com

Klein, G. A. (2015b). A naturalistic decision making perspective on studying intuitive decision making. *Journal of Applied Research in Memory and Cognition, 4*(3), 164–168. https://doi.org/10.1016/j.jarmac.2015.07.001

Klein, G. A., Borders, J., Newsome, E., Militello, L., & Klein, H. A. (2018). Cognitive skills training: Lessons learned. *Cognition, Technology & Work, 20*(4), 681–687. https://doi.org/10.1007/s10111-018-0528-5

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid Decision Making on the Fire Ground. *Proceedings of the Human Factors Society Annual Meeting*, *30*(6), 576–580. https://doi.org/10.1177/154193128603000616

Klein, G. A., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of Skilled Option Generation in Chess. *Organizational Behavior and Human Decision Processes*, *62*(1), 63–69. https://doi.org/10.1006/obhd.1995.1031

Kleinman, D., & Gollan, T. H. (2016). Speaking Two Languages for the Price of One: Bypassing Language Control Mechanisms via Accessibility-Driven Switches. *Psychological Science*, *27*(5), 700–714. https://doi.org/10.1177/0956797616634633

Klug, M., & Gramann, K. (2020). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *BioRxiv*, 2020.06.02.129213. https://doi.org/10.1101/2020.06.02.129213

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience & Biobehavioral Reviews*, *36*(1), 677–695. https://doi.org/10.1016/j.neubiorev.2011.10.002

Knyazeva, M. G., Barzegaran, E., Vildavski, V. Y., & Demonet, J.-F. (2018). Aging of human alpha rhythm. *Neurobiology of Aging*, *69*, 261–273. https://doi.org/10.1016/j.neurobiolaging.2018.05.018

Koda, T., & Maes, P. (1996). Agents with faces: The effect of personification. *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, 189–194. https://doi.org/10.1109/ROMAN.1996.568812

Kok, M., Hol, J. D., & Schön, T. B. (2017). Using inertial sensors for position and orientation estimation. *Foundations and Trends in Signal Processing*, *11*(1–2), 1–153. https://doi.org/10.1561/2000000094

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flügge, G., Korte, S. M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wöhr, M., & Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1291–1301. https://doi.org/10.1016/j.neubiorev.2011.02.003

Kothe, C., Medine, D., Boulay, C., Grivich, M., & Stenner. (2020). *Lab Streaming Layer* [HTML]. Swartz Center for Computational Neuroscience. https://github.com/sccn/labstreaminglayer (Original work published 2018)

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00417

Kuchera-Morin, J., Wright, M., Wakefield, G., Roberts, C., Adderton, D., Sajadi, B., Höllerer, T., & Majumder, A. (2014). Immersive full-surround multi-user system design. *Computers and Graphics (Pergamon)*, *40*(1), 10–21. https://doi.org/10.1016/j.cag.2013.12.004

Kuhl, J. (1992). A Theory of Self-regulation: Action versus State Orientation, Self-discrimination, and Some Applications. *Applied Psychology*, *41*(2), 97–129. https://doi.org/10.1111/j.1464-0597.1992.tb00688.x

Kvale, S. (1994). Ten standard Objections to Qualitative Research Interviews. *Journal of Phenomenological Psychology*, *25*(2), 147–173. https://doi.org/10.1163/156916294X00016

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16001837

Landman, A., Nieuwenhuys, A., & Oudejans, R. R. D. (2016a). Decision-related action orientation predicts police officers' shooting performance under pressure. *Anxiety, Stress and Coping*, *29*(5), 570–579. https://doi.org/10.1080/10615806.2015.1070834

Landman, A., Nieuwenhuys, A., & Oudejans, R. R. D. (2016b). The impact of personality traits and professional experience on police officers' shooting performance under pressure. *Ergonomics*, *59*(7), 950–961. https://doi.org/10.1080/00140139.2015.1107625

Landowska, A., Roberts, D., Eachus, P., & Barrett, A. (2018). Within- and Between-Session Prefrontal Cortex Response to Virtual Reality Exposure Therapy for Acrophobia. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00362

Landowska, A., Royle, S., Eachus, P., & Roberts, D. (2018). Testing the Potential of Combining Functional Near-Infrared Spectroscopy with Different Virtual Reality Displays—Oculus Rift and oCtAVE. In T. Jung & M. C. tom Dieck (Eds.), *Augmented Reality and Virtual Reality:*

Empowering Human, Place and Business (pp. 309–321). Springer International Publishing. https://doi.org/10.1007/978-3-319-64027-3_21

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*(3), 377–395.

Lanier, M., Waddell, T. F., Elson, M., Tamul, D. J., Ivory, J. D., & Przybylski, A. (2019). Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. *Computers in Human Behavior*, *100*, 70–78. https://doi.org/10.1016/j.chb.2019.06.015

Lavalle, S. M., Yershova, A., Katsev, M., & Antonov, M. (2014). Head tracking for the Oculus Rift. *Proceedings - IEEE International Conference on Robotics and Automation*, 187–194. https://doi.org/10.1109/ICRA.2014.6906608

Lawlor, P. (2016, June 29). *Keeping the virtual world stable in VR*. Qualcomm. https://www.qualcomm.com/news/onq/2016/06/29/keeping-virtual-world-stable-vr

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology*, *44*, 1–21. https://doi.org/10.1146/annurev.ps.44.020193.000245

Lenartowicz, A., & Poldrack, R. A. (2010). Brain Imaging. In G. F. Koob, M. L. Moal, & R. F. Thompson (Eds.), *Encyclopedia of Behavioral Neuroscience* (pp. 187–193). Academic Press. https://doi.org/10.1016/B978-0-08-045396-5.00052-X

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *JOSA A*, *22*(10), 2013–2033. https://doi.org/10.1364/JOSAA.22.002013

Leske, S., & Dalal, S. S. (2019). Reducing power line noise in EEG and MEG data via spectrum interpolation. *NeuroImage*, *189*, 763–776. https://doi.org/10.1016/j.neuroimage.2019.01.026

Li, B. J., & Bailenson, J. N. (2018). Exploring the Influence of Haptic and Olfactory Cues of a Virtual Donut on Satiation and Eating Behavior. *Presence*, *26*(03), 337–354. https://doi.org/10.1162/pres_a_00300

Lieberman, H., Faaborg, A., Daher, W., & Espinosa, J. (2005). How to wreck a nice beach you sing calm incense. *Proceedings of the 10th International Conference on Intelligent User Interfaces - IUI '05*, 278. https://doi.org/10.1145/1040830.1040898

Lighthall, N. R., Mather, M., & Gorlick, M. A. (2009). Acute stress increases sex differences in risk seeking in the Balloon Analogue Risk Task. *PLoS ONE*, *4*(7). https://doi.org/10.1371/journal.pone.0006002

Lighthall, N. R., Sakaki, M., Vasunilashorn, S., Nga, L., Somayajula, S., Chen, E. Y., Samii, N., & Mather, M. (2012). Gender differences in reward-related decision processing under stress. *Social Cognitive and Affective Neuroscience*, *7*(4), 476–484. https://doi.org/10.1093/scan/nsr026

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*(5), 331–352. https://doi.org/10.1002/bdm.381

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. https://doi.org/10.1006/obhd.1997.2679

Liu, H., Liang, L., Dunlap, S., Fan, N., & Chen, B. (2016). The effect of domain-general inhibition-related training on language switching: An ERP study. *Cognition*, *146*, 264–276. https://doi.org/10.1016/j.cognition.2015.10.004

Liu, Y., Dolan, R., Kurth-Nelson, Z., & Behrens, T. E. J. (2020). Measuring Sequences of Representations with Temporally Delayed Linear Modelling. *BioRxiv*, 2020.04.30.066407. https://doi.org/10.1101/2020.04.30.066407

Lockhart, T. S., Moore, R. A., Bard, K. A., & Stafford, L. D. (2019). Increases in theta CSD power and coherence during a calibrated stop-signal task: Implications for goal-conflict processing and the Behavioural Inhibition System. *Personality Neuroscience*, *2*. https://doi.org/10.1017/pen.2019.10

Lombard, M., & Jones, M. T. (2015). Defining Presence. In M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, & R. J. Schaevitz (Eds.), *Immersed in Media: Telepresence Theory, Measurement & Technology* (pp. 13–34). Springer International Publishing. https://doi.org/10.1007/978-3-319-10190-3_2

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, and Computers*, *31*(4), 557–564. https://doi.org/10.3758/BF03200735

Lucassen, P. J., Pruessner, J., Sousa, N., Almeida, O. F. X., Van Dam, A. M., Rajkowska, G., Swaab, D. F., & Czéh, B. (2014). Neuropathology of stress. *Acta Neuropathologica*, *127*(1), 109–135. https://doi.org/10.1007/s00401-013-1223-5

Luce, R. D., & von Winterfeldt, D. (1994). What Common Ground Exists for Descriptive, Prescriptive, and Normative Utility Theories? *Management Science*, *40*(2), 263–279. https://doi.org/10.1287/mnsc.40.2.263

Ludwig, C. J. H., Alexander, N., Howard, K. L., Jedrzejewska, A. A., Mundkur, I., & Redmill, D. (2018). The influence of visual flow and perceptual load on locomotion speed. *Attention, Perception, and Psychophysics*, *80*(1), 69–81. https://doi.org/10.3758/s13414-017-1417-3

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, *115*(8), 1821–1835. https://doi.org/10.1016/j.clinph.2004.03.031

Maguire, E. A., Henson, R. N. A., Mummery, C. J., & Frith, C. D. (2001). Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *NeuroReport*, *12*(3), 441–444.

Mahjoory, K., Nikulin, V. V., Botrel, L., Linkenkaer-Hansen, K., Fato, M. M., & Haufe, S. (2017). Consistency of EEG source localization and connectivity estimates. *NeuroImage*, *152*(August 2016), 590–601. https://doi.org/10.1016/j.neuroimage.2017.02.076

Mangels, L., Suss, J., & Lande, B. (2020). Police Expertise and Use of Force: Using a Mixed-Methods Approach to Model Expert and Novice Use-of-Force Decision-Making. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-020-09364-4

Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience*, *10*. https://doi.org/10.3389/fncom.2016.00094

Marewski, J. N., & Schooler, L. J. (2011). Cognitive Niches: An Ecological Model of Strategy Selection. *Psychological Review*, *118*(3), 393–437. https://doi.org/10.1037/a0024143

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marschak, J. (1950). Rational Behavior, Uncertain Prospects, and Measurable Utility. *Econometrica*, *18*(2), 111. https://doi.org/10.2307/1907264

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real: Physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology*, *33*(10), 1264–1273. https://doi.org/10.1177/0269881119860156

Mason, J. W. (1968a). A review of psychoendocrine research on the pituitary-adrenal cortical system. *Psychosomatic Medicine*, *30*(5), 576–607. https://doi.org/10.1097/00006842-196809000-00020

Mason, J. W. (1968b). A review of psychoendocrine research on the sympathetic-adrenal medullary system. *Psychosomatic Medicine*, *30*(5), 631–653. https://doi.org/10.1097/00006842-196809000-00022

Matamala-Gomez, M., Nierula, B., Donegan, T., Slater, M., & Sanchez-Vives, M. V. (2020). Manipulating the Perceived Shape and Color of a Virtual Limb Can Modulate Pain Responses. *Journal of Clinical Medicine*, *9*(2), 291. https://doi.org/10.3390/jcm9020291

Medvedovsky, M., Taulu, S., Bikmullina, R., & Paetau, R. (2007). Artifact and head movement compensation in MEG. *Neurology, Neurophysiology, and Neuroscience*, 4.

Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage*, *61*(2), 371–385. https://doi.org/10.1016/j.neuroimage.2011.12.039

Militello, L., Lipshitz, R., & Schraagen, J. M. (2017). *Naturalistic Decision Making and Macrocognition*. CRC Press, Boca Raton, FL, USA.

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology*, *14*(2), 69–97. https://doi.org/10.1016/0301-0082(80)90018-0

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement*, *33*(2), 259–270. https://doi.org/10.1088/0967-3334/33/2/259

Mori, M. (1970). Bukimi no tani. *Energy*, *7*(4), 33–35.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley. *IEEE Robotics Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Morrison, G. B., & Vila, B. J. (1998). Police handgun qualification: Practical measure or aimless activity? *Policing: An International Journal of Police Strategies & Management*, *21*(3), 510–533. https://doi.org/10.1108/13639519810228804

Muhanna, M. A. (2015). Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences*, *27*(3), 344–361. https://doi.org/10.1016/j.jksuci.2014.03.023

Muñoz, J. E., Quintero, L., Stephens, C. L., & Pope, A. T. (2020). A Psychophysiological Model of Firearms Training in Police Officers: A Virtual Reality Experiment for Biocybernetic Adaptation. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00683

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Frontiers in Human Neuroscience*, *7*(MAR), 1–11. https://doi.org/10.3389/fnhum.2013.00138

Naour, T. L., Courty, N., & Gibet, S. (2019). Skeletal mesh animation driven by few positional constraints. *Computer Animation and Virtual Worlds*, *30*(3–4), e1900. https://doi.org/10.1002/cav.1900

Nathan, K., & Contreras-Vidal, J. L. (2016). Negligible Motion Artifacts in Scalp Electroencephalography (EEG) During Treadmill Walking. *Frontiers in Human Neuroscience*, *9*, 708. https://doi.org/10.3389/fnhum.2015.00708

National Policing Improvement Agency. (2011). *Manual of Guidance on the Management, Command and Deployment of Armed Officers* (pp. 1–140). National Policing Improvement Agency.

Nauta, W. J., & Mehler, W. R. (1966). Projections of the lentiform nucleus in the monkey. *Brain Research*, *1*(1), 3–42. https://doi.org/10.1016/0006-8993(66)90103-x

Nenna, F., Do, C.-T., Protzak, J., & Gramann, K. (2020). Alteration of brain dynamics during natural dual-task walking. *BioRxiv*, 2020.02.27.968164. https://doi.org/10.1101/2020.02.27.968164

Nibbeling, N., Oudejans, R. R. D., Ubink, E. M., & Daanen, H. A. M. (2014). The effects of anxiety and exercise-induced fatigue on shooting accuracy and cognitive performance in infantry soldiers. *Ergonomics*, *57*(9), 1366–1379. https://doi.org/10.1080/00140139.2014.924572

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective and Behavioral Neuroscience*, *3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17

Nieuwenhuys, A. (2011). *Comparison of step-fire and fire-step shooting strategies: Police officers' shot accuracy during annual shooting tests*. Unpublished.

Nieuwenhuys, A., Cañal-Bruland, R., & Oudejans, R. R. D. (2012). Effects of Threat on Police Officers' Shooting Behavior: Anxiety, Action Specificity, and Affective Influences on Perception. *Applied Cognitive Psychology*, *26*(4), 608–615. https://doi.org/10.1002/acp.2838

Nieuwenhuys, A., Oudejans, R. R. D., & Savelsbergh, G. J. P. (2012). Shoot or don't shoot? Why police officers are more inclined to shoot when they are anxious. *Emotion*, *12*(4), 827–833. https://doi.org/10.1037/a0025699

Nieuwenhuys, A., Weber, J., van der Hoeve, R., & Oudejans, R. R. D. (2017). Sitting duck or scaredy-cat? Effects of shot execution strategy on anxiety and police officers' shooting performance under high threat. *Legal and Criminological Psychology*, *22*(2), 274–287. https://doi.org/10.1111/lcrp.12099

Nilsson, D., & Kinateder, M. (2015). Virtual Reality Experiments—The future or a dead end? *6th Human Behaviour in Fire Symposium*, 13–22.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., … Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

O'Connor, A. M., Rostom, A., Fiset, V., Tetroe, J., Entwistle, V., Llewellyn-Thomas, H., Holmes-Rovner, M., Barry, M., & Jones, J. (1999). Decision aids for patients facing health treatment or screening decisions: Systematic review. *BMJ*, *319*(7212), 731–734. https://doi.org/10.1136/bmj.319.7212.731

Office of the High Commissioner. (1990). *OHCHR | Basic Principles on the Use of Force and Firearms*. https://www.ohchr.org/EN/ProfessionalInterest/Pages/UseOfForceAndFirearms.aspx

Oh, C., Herrera, F., & Bailenson, J. (2019). The Effects of Immersion and Real-World Distractions on Virtual Social Interactions. *Cyberpsychology, Behavior, and Social Networking*, *22*(6), 365–372. https://doi.org/10.1089/cyber.2018.0404

Öhman, A., & Dimberg, U. (1978). Facial expressions as conditioned stimuli for electrodermal responses: A case of ‘preparedness’? *Journal of Personality and Social Psychology*, *36*(11), 1251–1258. https://doi.org/10.1037/0022-3514.36.11.1251

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The Face in the Crowd Revisited: A Threat Advantage With Schematic Stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381–396.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, *47*(6), 419–427. https://doi.org/10.1037/h0058775

Oostendorp, T. F., van Oosterom, A., & Huiskamp, G. (1989). Interpolation on a triangulated 3D surface. *Journal of Computational Physics*, *80*(2), 331–343. https://doi.org/10.1016/0021-9991(89)90103-4

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Oostenveld, R., & Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human Brain Mapping*, *17*(3), 179–192. https://doi.org/10.1002/hbm.10061

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*(5), 529–541. https://doi.org/10.1109/PROC.1981.12022

O’Regan, J. K. (1992). Solving the ‘real’ mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488. https://doi.org/10.1037/h0084327

Oudejans, R. R. D. (2008). Reality-based practice under pressure improves handgun shooting performance of police officers. *Ergonomics*, *51*(3), 261–273. https://doi.org/10.1080/00140130701577435

pally qle. (2016, September). *Unreal Engine 4 Good to Know—Read and Write Text File* [YouTube]. https://www.youtube.com/watch?v=wemG-YI2iag

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Papazoglou, K., & Andersen, J. P. (2014). A guide to utilizing police training as a tool to promote resilience and improve health outcomes among police officers. *Traumatology*, *20*(2), 103–111. https://doi.org/10.1037/h0099394

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*(DEC), 1–19. https://doi.org/10.3389/fnhum.2015.00660

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*(1), 49–65. https://doi.org/10.1016/0167-8760(84)90014-X

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive Strategy Selection in Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, *22*(3), 779–787. https://doi.org/10.1016/j.concog.2013.04.016

Peeters, D. (2020). Bilingual switching between languages and listeners: Insights from immersive virtual reality. *Cognition*, *195*(October 2019), 104107. https://doi.org/10.1016/j.cognition.2019.104107

Peeters, D., & Dijkstra, T. (2018). Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Bilingualism: Language and Cognition*, *21*(5), 1035–1061. https://doi.org/10.1017/S1366728917000396

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pessoa, L. (2011). Reprint of: Emotion and cognition and the amygdala: From ‘ what is it?’ to ‘ what’s to be done?’ *Neuropsychologia*, *49*(4), 681–694. https://doi.org/10.1016/j.neuropsychologia.2011.02.030

Petras, K., Oever, S. T., & Jansma, B. M. (2016). The effect of distance on moral engagement: Event related potentials and alpha power are sensitive to perspective in a virtual shooting task. *Frontiers in Psychology*, *6*(JAN), 1–9. https://doi.org/10.3389/fpsyg.2015.02008

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, *110*(11), 1842–1857. https://doi.org/10.1016/S1388-2457(99)00141-8

Phillips, C., Rugg, M. D., & Friston, K. J. (2000). Direct extraction of realistic constraints from T1 structural MR image for electromagnetic source localisation. *NeuroImage*, *11*(5 PART II), 5491. https://doi.org/10.1016/s1053-8119(00)91422-8

Pierson, W. R., & Montoye, H. J. (1958). Movement Time, Reaction Time and Age. *Journal of Gerontology*, *13*(4), 418–421. https://doi.org/10.1093/geronj/13.4.418

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Plant, E. A., & Peruche, B. M. (2005). The Consequences of Race for Police Officers’ Responses to Criminal Suspects. *Psychological Science*, *16*(3), 180–183. https://doi.org/10.1111/j.0956-7976.2005.00800.x

Polich, J. (1997). EEG and ERP assessment of normal aging. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *104*(3), 244–256. https://doi.org/10.1016/S0168-5597(97)96139-6

Pollick, F. E. (2010). In Search of the Uncanny Valley. In P. Daras & O. M. Ibarra (Eds.), *User Centric Media* (pp. 69–78). Springer. https://doi.org/10.1007/978-3-642-12630-7_8

Popa, L., Selejan, O., Scott, A., Mureşanu, D. F., Balea, M., & Rafila, A. (2015). Reading beyond the glance: Eye tracking in neurosciences. *Neurological Sciences: Official Journal of the Italian*

Neurological Society and of the Italian Society of Clinical Neurophysiology, *36*(5), 683–688. https://doi.org/10.1007/s10072-015-2076-6

Porcelli, A. J. (2014). An alternative to the traditional cold pressor test: The cold pressor arm wrap. *Journal of Visualized Experiments*, *83*(83), e50849. https://doi.org/10.3791/50849

Porcelli, A. J., & Delgado, M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychological Science*, *20*(3), 278–283. https://doi.org/10.1111/j.1467-9280.2009.02288.x

Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience*, *13*(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Poulton, E. C. (1957). On prediction in skilled movements. *Psychological Bulletin*, *54*(6), 467–478. http://dx.doi.org/10.1037/h0045515

Preston, S. D., Buchanan, T. W., Stansfield, R. B., & Bechara, A. (2007). Effects of Anticipatory Stress on Decision Making in a Gambling Task. *Behavioral Neuroscience*, *121*(2), 257–263. https://doi.org/10.1037/0735-7044.121.2.257

Prokop, T., Schubert, M., & Berger, W. (1997). Visual influence on human locomotion. Modulation to changes in optic flow. *Experimental Brain Research*, *114*(1), 63–70. https://doi.org/10.1007/PL00005624

Pruessner, J. C., Dedovic, K., Pruessner, M., Lord, C., Buss, C., Collins, L., Dagher, A., & Lupien, S. J. (2010). Stress regulation in the central nervous system: Evidence from structural and functional neuroimaging studies in human populations—2008 Curt Richter Award Winner. *Psychoneuroendocrinology*, *35*(1), 179–191. https://doi.org/10.1016/j.psyneuen.2009.02.016

Puce, A., & Hämäläinen, M. S. (2017). A Review of Issues Related to Data Acquisition and Analysis in EEG/MEG Studies. *Brain Sciences*, *7*(6). https://doi.org/10.3390/brainsci7060058

Qin, L., Gelderen, P. van, Derbyshire, J. A., Jin, F., Lee, J., Zwart, J. A. de, Tao, Y., & Duyn, J. H. (2009). Prospective head-movement correction for high-resolution MRI using an in-bore optical tracking system. *Magnetic Resonance in Medicine*, *62*(4), 924–934. https://doi.org/10.1002/mrm.22076

Quaresima, V., & Ferrari, M. (2019). Functional Near-Infrared Spectroscopy (fNIRS) for Assessing Cerebral Cortex Function During Human Behavior in Natural/Social Situations: A Concise

Review. *Organizational Research Methods*, *22*(1), 46–68. https://doi.org/10.1177/1094428116658959

Queiros, C., Kaiseler, M., & Leitão da Silva, A. (2015). Burnout as predictor of aggressivity among police officers. *European Journal of Policing Studies*, *1*(2), 110–134.

R Development Core Team. (2019). *R: A Language and Environment for Statistical Computing* (3.6.1) [Computer software]. http://www.r-project.org/

Reader, A. T., & Holmes, N. P. (2016). Examining ecological validity in social interaction: Problems of visual fidelity, gaze, and social potential. *Culture and Brain*, *4*(2), 134–146. https://doi.org/10.1007/s40167-016-0041-8

Reaves, B. A. (2009). State and Local Law Enforcement Training Academies, 2006. *Bureau of Justice Statistics*, 15.

Reaves, B. A. (2016). State and Local Law Enforcement Training Academies, 2013. *Bureau of Justice Statistics*, 20.

Redcay, E., & Moraczewski, D. (2019). Social cognition in context: A naturalistic imaging approach. *NeuroImage*, 116392. https://doi.org/10.1016/j.neuroimage.2019.116392

Renden, P. G., Landman, A., Daalder, N. R., de Cock, H. P., Savelsbergh, G. J. P., & Oudejans, R. R. D. (2017). Effects of threat, trait anxiety and state anxiety on police officers' actions during an arrest. *Legal and Criminological Psychology*, *22*(1), 116–129. https://doi.org/10.1111/lcrp.12077

Renden, P. G., Landman, A., Geerts, S. F., Jansen, S. E. M., Faber, G. S., Savelsbergh, G. J. P., & Oudejans, R. R. D. (2014). Effects of anxiety on the execution of police arrest and self-defense skills. *Anxiety, Stress and Coping*, *27*(1), 100–112. https://doi.org/10.1080/10615806.2013.810213

Ries, A. J., Touryan, J., Vettel, J., McDowell, K., & Hairston, W. D. (2014). A Comparison of Electroencephalography Signals Acquired from Conventional and Mobile Systems. *Journal of Neuroscience and Neuroengineering*, *3*(1), 10–20. https://doi.org/10.1166/jnsne.2014.1092

Roberts, D., Heldal, I., Otto, O., & Wolff, R. (2006). Factors influencing flow of object focussed collaboration in collaborative virtual environments. *Virtual Reality*, *10*(2), 119–133. https://doi.org/10.1007/s10055-006-0050-6

Roberts, G., Holmes, N., Alexander, N., Boto, E., Leggett, J., Hill, R. M., Shah, V., Rea, M., Vaughan, R., Maguire, E. A., Kessler, K., Beebe, S., Fromhold, M., Barnes, G. R., Bowtell, R., & Brookes, M. J. (2019). Towards OPM-MEG in a virtual reality environment. *NeuroImage*, *199*(June), 408–417. https://doi.org/10.1016/j.neuroimage.2019.06.010

Rocchini, C., Cignoni, P., Montani, C., Pingi, P., & Scopigno, R. (2001). A low cost 3D scanner based on structured light. *Computer Graphics Forum*, *20*(3), 299–308. https://doi.org/10.1111/1467-8659.00522

Rogers, M. D. (2003). Police Force! An Examination of the Use of Force, Firearms and Less-Lethal Weapons by British Police. *The Police Journal: Theory, Practice and Principles*, *76*(3), 189–203. https://doi.org/10.1350/pojo.76.3.189.19443

Rosenberg, R. S., Baughman, S. L., & Bailenson, J. N. (2013). Virtual Superheroes: Using Superpowers in Virtual Reality to Encourage Prosocial Behavior. *PLoS ONE*, *8*(1), 1–9. https://doi.org/10.1371/journal.pone.0055003

Ross, K. G., Klein, G. A., Thunholm, P., Schmitt, J. F., & Baxter, H. C. (2004). *The Recognition-Primed Decision Model*. ARMY COMBINED ARMS CENTER FORT LEAVENWORTH KS MILITARY REVIEW. https://apps.dtic.mil/docs/citations/ADA521492

Roy, B. (1990). Decision-aid and decision-making. *European Journal of Operational Research*, *45*(2), 324–331. https://doi.org/10.1016/0377-2217(90)90196-I

Roy, B. (1993). Decision science or decision-aid science? *European Journal of Operational Research*, *66*(2), 184–203. https://doi.org/10.1016/0377-2217(93)90312-B

Rugg, M. D., & Henson, R. N. (2005). Episodic memory retrieval: An (eventrelated) functional neuroimaging perspective. *The Cognitive Neuroscience of Memory Encoding and Retrieval*, 707–714.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience*, *8*(12), 1647–1650. https://doi.org/10.1038/nn1606

Saga, Y., Richard, A., Sgambato-Faure, V., Hoshi, E., Tobler, P. N., & Tremblay, L. (2017). Ventral Pallidum Encodes Contextual Information and Controls Aversive Behaviors. *Cerebral Cortex*, *27*(4), 2528–2543. https://doi.org/10.1093/cercor/bhw107

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332–339. https://doi.org/10.1038/nrn1651

Savage, L. J. (1954). *The foundations of statistics*. John Wiley & Sons, Oxford, EK.

Schäfer, M., Rübsamen, R., Dörrscheidt, G. J., & Knipschild, M. (1992). Setting complex tasks to single units in the avian auditory forebrain. II: Do we really need natural stimuli to describe neuronal response characteristics? *Hearing Research*, *57*(2), 231–244. https://doi.org/10.1016/0378-5955(92)90154-F

Schecklmann, M., Mann, A., Langguth, B., Ehlis, A. C., Fallgatter, A. J., & Haeussinger, F. B. (2017). The temporal muscle of the head can cause artifacts in optical imaging studies with functional near-infrared spectroscopy. *Frontiers in Human Neuroscience*, *11*(September), 1–13. https://doi.org/10.3389/fnhum.2017.00456

Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *65*(5), 344–360. https://doi.org/10.1016/0168-5597(86)90014-6

Schiza, E., Matsangidou, M., Neokleous, K., & Pattichis, C. S. (2019). Virtual Reality Applications for Neurological Disease: A Review. *Frontiers in Robotics and AI*, *6*. https://doi.org/10.3389/frobt.2019.00100

Schmiedt-Fehr, C., Dühl, S., & Basar-Eroglu, C. (2011). Age-related increases in within-person variability: Delta and theta oscillations indicate that the elderly are not always old. *Neuroscience Letters*, *495*(2), 159–163. https://doi.org/10.1016/j.neulet.2011.03.062

Schoemaker, P. J. H. (1982). The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature*, *20*(2), 529–563. JSTOR.

Schöne, B., Wessels, M., & Gruber, T. (2019). Experiences in Virtual Reality: A Window to Autobiographical Memory. *Current Psychology*, *38*(3), 715–719. https://doi.org/10.1007/s12144-017-9648-y

Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, *194*(3), 465–475. https://doi.org/10.1007/s00221-009-1721-9

Schwartz, M. D., Valdimarsdottir, H. B., DeMarco, T. A., Peshkin, B. N., Lawrence, W., Rispoli, J., Brown, K., Isaacs, C., O'Neill, S., Shelby, R., Grumet, S. C., McGovern, M. M., Garnett, S., Bremer, H., Leaman, S., O'Mara, K., Kelleher, S., & Komaridis, K. (2009). Randomized trial of a decision aid for BRCA1/BRCA2 mutation carriers: Impact on measures of decision making and satisfaction. *Health Psychology*, *28*(1), 11–19. https://doi.org/10.1037/a0013147

Seligman, M. (1970). On the generality of the laws of learning. *Psychological Review*, *77*(5), 406–418.

Selye, H. (1936). A Syndrome produced by Diverse Nocuous Agents. *Nature*, *138*(3479), 32–32. https://doi.org/10.1038/138032a0

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., & Wüst, S. (2016). Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology*, *110*, 47–55. https://doi.org/10.1016/j.ijpsycho.2016.10.008

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, *11*(8), 11–11. https://doi.org/10.1167/11.8.11

Shiv, B., Loewenstein, G., & Bechara, A. (2005). The dark side of emotion in decision-making: When individuals with decreased emotional reactions make more advantageous decisions. *Cognitive Brain Research*, *23*(1), 85–92. https://doi.org/10.1016/j.cogbrainres.2005.01.006

Siegel, Z. D., & Kelly, J. W. (2017). Walking through a virtual environment improves perceived size within and beyond the walked space. *Attention, Perception, and Psychophysics*, *79*(1), 39–44. https://doi.org/10.3758/s13414-016-1243-z

Silva, A. R., Pinho, M. S., Macedo, L., & Moulin, C. J. A. (2018). A critical review of the effects of wearable cameras on memory. *Neuropsychological Rehabilitation*, *28*(1), 117–141. https://doi.org/10.1080/09602011.2015.1128450

Simon, H. A. (1990). Bounded Rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Utility and Probability* (pp. 15–18). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20568-4_5

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *8*(5), 560–565. https://doi.org/10.1162/105474699566477

Slater, M. (2004). How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *13*(4), 484–493. https://doi.org/10.1162/1054746041944849

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*(3), 431–433. https://doi.org/10.1111/bjop.12305

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., Watson, Z., Breen, G., Schwarz, R., Steptoe, W., Szostak, D., Halan, S., Fox, D., & Silver, J. (2020). The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality*, *1*. https://doi.org/10.3389/frvir.2020.00001

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoperators and Virtual Environments*, *9*(1), 37–51. https://doi.org/10.1162/105474600566600

Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater, M., Usoh, M., & Steed, A. (1995). Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, *2*(3), 201–219. https://doi.org/10.1145/210079.210084

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Slooter, J., & Norren, D. van. (1980). Visual acuity measured with pupil responses to checkerboard stimuli. *Investigative Ophthalmology & Visual Science*, *19*(1), 105–108.

Smeets, T., Cornelisse, S., Quaedflieg, C. W. E. M., Meyer, T., Jelicic, M., & Merckelbach, H. (2012). Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology*, *37*(12), 1998–2008. https://doi.org/10.1016/j.psyneuen.2012.04.012

Smith, E. E., Tenke, C. E., Deldin, P. J., Trivedi, M. H., Weissman, M. M., Auerbach, R. P., Bruder, G. E., Pizzagalli, D. A., & Kayser, J. (2020). Frontal theta and posterior alpha in resting EEG:

A critical examination of convergent and discriminant validity. *Psychophysiology, 57*(2), e13483. https://doi.org/10.1111/psyp.13483

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Spierer, L., Chavan, C., & Manuel, A. L. (2013). Training-induced behavioral and brain plasticity in inhibitory control. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00427

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stacey, D., Légaré, F., Lewis, K., Barry, M. J., Bennett, C. L., Eden, K. B., Holmes-Rovner, M., Llewellyn-Thomas, H., Lyddiatt, A., Thomson, R., & Trevena, L. (2017). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, *4*. https://doi.org/10.1002/14651858.CD001431.pub5

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–726. https://doi.org/10.1017/S0140525X00003435

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1228–1248. https://doi.org/10.1016/j.neubiorev.2012.02.003

Starcke, K., & Brand, M. (2016). Effects of stress on decisions under uncertainty: A meta-analysis. *Psychological Bulletin*, *142*(9), 909–933. https://doi.org/10.1037/bul0000060

Starcke, K., Wolf, O. T., Markowitsch, H. J., & Brand, M. (2008). Anticipatory Stress Influences Decision Making Under Explicit Risk Conditions. *Behavioral Neuroscience*, *122*(6), 1352–1360. https://doi.org/10.1037/a0013281

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Strange, D., & Takarangi, M. K. T. (2015). Memory Distortion for Traumatic Events: The Role of Mental Imagery. *Frontiers in Psychiatry*, *6*. https://doi.org/10.3389/fpsyt.2015.00027

Suss, J., & Boulton, L. (2019). Expertise in Law Enforcement. In *The Oxford Handbook of Expertise* (1st ed., pp. 765–791). Oxford University Press.

Suss, J., & Ward, P. (2018). Revealing perceptual–cognitive expertise in law enforcement: An iterative approach using verbal-report, temporal-occlusion, and option-generation methods. *Cognition, Technology and Work*, *20*(4), 585–596. https://doi.org/10.1007/s10111-018-0493-z

Sutherland, I. E. (1965). The ultimate display. *Proceedings of the International Federation of Information Processing Congress*, *2*, 506–508.

Sven. (2019). *inpolyhedron—Are points inside a triangulated volume?* (3.3.0.0) [MATLAB]. https://www.mathworks.com/matlabcentral/fileexchange/37856-inpolyhedron-are-points-inside-a-triangulated-volume

Tadel, F., Bock, E., Niso, G., Mosher, J. C., Cousineau, M., Pantazis, D., Leahy, R. M., & Baillet, S. (2019). MEG/EEG Group Analysis With Brainstorm. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00076

Tajima, D., Mizuno, T., Kume, Y., & Yoshida, T. (2015). The mirror illusion: Does proprioceptive drift go hand in hand with sense of agency? *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00200

Taylor, P. L. (2019). Dispatch Priming and the Police Decision to Use Deadly Force. *Police Quarterly*, 1098611119896653. https://doi.org/10.1177/1098611119896653

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*(2), 1–11.

Török, Á., Sulykos, I., Kecskés-Kovács, K., Persa, G., Galambos, P., Kóbor, A., Czigler, I., Csépe, V., Baranyi, P., & Honbolygó, F. (2014). Comparison between wireless and wired EEG recordings in a virtual reality lab: Case report. *5th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2014 - Proceedings*, 599–603. https://doi.org/10.1109/CogInfoCom.2014.7020414

Tremmel, C., Herff, C., & Krusienski, D. J. (2019). EEG Movement Artifact Suppression in Interactive Virtual Reality. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4576–4579. https://doi.org/10.1109/EMBC.2019.8856961

Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods*, *50*(2), 862–869. https://doi.org/10.3758/s13428-017-0911-9

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301–306. https://doi.org/10.1126/science.2296719

Tversky, A., & Kahneman, D. (1986). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, Cambridge, UK.

Tversky, A., & Kahneman, D. (2017). Rational choice and the framing of decisions. *Decision Science*, *59*(4), 43–70. https://doi.org/10.1086/296365

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503. https://doi.org/10.1162/105474600566989

Valmaggia, L. R., Latif, L., Kempton, M. J., & Rus-Calafell, M. (2016). Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. *Psychiatry Research*, *236*, 189–195. https://doi.org/10.1016/j.psychres.2016.01.015

Van de Steen, F., Faes, L., Karahan, E., Songsiri, J., Valdes-Sosa, P. A., & Marinazzo, D. (2019). Critical Comments on EEG Sensor Space Dynamical Connectivity Analysis. *Brain Topography*, *32*(4), 643–654. https://doi.org/10.1007/s10548-016-0538-7

van den Bos, R., Harteveld, M., & Stoop, H. (2009). Stress and decision-making in humans: Performance is related to cortisol reactivity, albeit differently in men and women. *Psychoneuroendocrinology*, *34*(10), 1449–1458. https://doi.org/10.1016/j.psyneuen.2009.04.016

van den Bos, R., Taris, R., Scheppink, B., de Haan, L., & Verster, J. C. (2014). Salivary cortisol and alpha-amylase levels during an assessment procedure correlate differently with risk-taking measures in male and female police recruits. *Frontiers in Behavioral Neuroscience*, *7*(JAN), 219. https://doi.org/10.3389/fnbeh.2013.00219

van den Broek, S. P., Reinders, F., Donderwinkel, M., & Peters, M. J. (1998). Volume conduction effects in EEG and MEG. *Electroencephalography and Clinical Neurophysiology*, *106*(6), 522–534. https://doi.org/10.1016/S0013-4694(97)00147-8

Vasconcelos-Raposo, J., Bessa, M., Melo, M., Barbosa, L., Rodrigues, R., Teixeira, C. M., Cabral, L., & Sousa, A. A. (2016). Adaptation and Validation of the Igroup Presence Questionnaire (IPQ) in a Portuguese Sample. *Presence: Teleoperators and Virtual Environments*, *25*(3), 191–203. https://doi.org/10.1162/PRES_a_00261

von Neumann, J., & Morgenstern, O. (2007). Theory of games and economic behavior. In *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, US. https://doi.org/10.2307/3610940

Wachowiak, M. (2011). All in a Sniff: Olfaction as a Model for Active Sensing. *Neuron*, *71*(6), 962–973. https://doi.org/10.1016/j.neuron.2011.08.030

Wade, A. R., McCall, C., Karapanagiotidis, T., Schofield, G., Preston, C., Hartley, T., Kaestner, M., Horner, A., Maloney, R. T., Smallwood, J., Jefferies, E., Bloj, M., & Harris, J. (2018). A neuroscientific approach to exploring fundamental questions in VR. *IS and T International Symposium on Electronic Imaging Science and Technology*, *3*, 1–6. https://doi.org/10.2352/ISSN.2470-1173.2018.03.ERVR-435

Wade, K. A., Nash, R. A., & Lindsay, D. S. (2018). Reasons to Doubt the Reliability of Eyewitness Memory: Commentary on Wixted, Mickes, and Fisher (2018). *Perspectives on Psychological Science*, *13*(3), 339–342. https://doi.org/10.1177/1745691618758261

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wang, B. A., Viswanathan, S., Abdollahi, R. O., Rosjat, N., Popovych, S., Daun, S., Grefkes, C., & Fink, G. R. (2017). Frequency-specific modulation of connectivity in the ipsilateral sensorimotor cortex by different forms of movement initiation. *NeuroImage*, *159*, 248–260. https://doi.org/10.1016/j.neuroimage.2017.07.054

Wang, C.-H., Chang, C.-C., Liang, Y.-M., Shih, C.-M., Chiu, W.-S., Tseng, P., Hung, D. L., Tzeng, O. J. L., Muggleton, N. G., & Juan, C.-H. (2013). Open vs. Closed Skill Sports and the Modulation of Inhibitory Control. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0055773

Wang, Z., & Simoncelli, E. P. (2004). Local Phase Coherence and the Perception of Blur. *Advances in Neural Information Processing Systems*, 1435–1442.

Ward, P., Suss, J., Eccles, D. W., Williams, A. M., & Harris, K. R. (2011). Skill-based differences in option generation in a complex task: A verbal protocol analysis. *Cognitive Processing*, *12*(3), 289–300. https://doi.org/10.1007/s10339-011-0397-9

Weber, M., & Camerer, C. (1987). Recent developments in modelling preferences under risk. *Operations-Research-Spektrum*, *9*(3), 129–151. https://doi.org/10.1007/BF01721094

Weber, S., Mast, F. W., & Weibel, D. (2019). Body size illusions influence perceived size of objects: A validation of previous research in virtual reality. *Virtual Reality*. https://doi.org/10.1007/s10055-019-00402-z

Webster, M. A., Georgeson, M. A., & Webster, S. M. (2002). Neural adjustments to image blur. *Nature Neuroscience*, *5*(9), 839–840. https://doi.org/10.1038/nn906

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00158

Weller, J. A., Levin, I. P., Shiv, B., & Bechara, A. (2007). Neural correlates of adaptive decision making for risky gains and losses. *Psychological Science*, *18*(11), 958–964. https://doi.org/10.1111/j.1467-9280.2007.02009.x

West, B. J., & Shlesinger, M. (1990). The Noise in Natural Phenomena. *American Scientist*, *78*(1), 40–45. JSTOR.

Wiesing, M., Fink, G. R., & Weidner R. (2019, October 20). *Accuracy and Precision of Stimulus Timing and Reaction Times with Unreal Engine and SteamVR* [Poster]. Society for Neuroscience 2019, Chicago, IL, USA. https://www.abstractsonline.com/pp8/#!/7883/presentation/52893

Wilhelms, J. (1987). Using Dynamic Analysis for Realistic Animation of Articulated Bodies. *IEEE Computer Graphics and Applications*, *7*(6), 12–27. https://doi.org/10.1109/MCG.1987.276893

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, *34*(9), 1187–1195. https://doi.org/10.1016/0042-6989(94)90300-X

Wood, C. C., & Jennings, J. R. (1976). Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception & Psychophysics*, *19*(1), 92–102. https://doi.org/10.3758/BF03199392

Worrall, J. L., Bishopp, S. A., Zinser, S. C., Wheeler, A. P., & Phillips, S. W. (2018). Exploring Bias in Police Shooting Decisions With Real Shoot/Don't Shoot Cases. *Crime & Delinquency*, *64*(9), 1171–1192. https://doi.org/10.1177/0011128718756038

Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin & Review*, *20*(2), 364–371. https://doi.org/10.3758/s13423-012-0324-9

Wunderlich, A., & Gramann, K. (2020). Brain dynamics of assisted pedestrian navigation in the real-world. *BioRxiv*, 2020.06.08.139469. https://doi.org/10.1101/2020.06.08.139469

Yates, J. F., Veinott, E. S., & Patalano, A. L. (2003). Hard Decisions, Bad Decisions: On Decision Quality and Decision Aiding. In S. L. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (1st ed., pp. 13–63). Cambridge University Press. https://doi.org/10.1017/CBO9780511609978.003

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

You, Y., Ma, Y., Ji, Z., Meng, F., Li, A., & Zhang, C. (2018). Unconscious response inhibition differences between table tennis athletes and non-athletes. *PeerJ*, *2018*(9), e5548. https://doi.org/10.7717/peerj.5548

Yuan, Y., & Steed, A. (2010). Is the rubber hand illusion induced by immersive virtual reality? *2010 IEEE Virtual Reality Conference (VR)*, 95–102. https://doi.org/10.1109/VR.2010.5444807

Zeelenberg, M., Nelissen, R. M. A., Breugelmans, S. M., & Pieters, R. (2008). On emotion specificity in decision making: Why feeling is for doing. *Judgment and Decision Making*, *3*(1), 18–27.

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., & Domes, G. (2019). Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*, *101*, 186–192. https://doi.org/10.1016/j.psyneuen.2018.11.010

Zimmermann, R., & Scharein, E. (2004). MEG and EEG show different sensitivity to myogenic artifacts. *Neurology & Clinical Neurophysiology : NCN*, *2004*, 78–78.