Measuring the Decision Accuracy and Decision Confidence of Air Defence Operators

Thesis submitted in accordance with the requirements of The University of Liverpool for the degree of Doctor in Philosophy

By

Jade Elizabeth Adams-White

January 2019

ABSTRACT

Air defence decision making is demanding, time pressured and complex. Operators must make complex and cognitively demanding decisions in dynamic and uncertain environments. This thesis sought to increase the understanding of, and measure, decision confidence and accuracy in air defence decision making. In doing so, a novel method was designed and developed which was based on an integration of Classical Decision Making (CDM) and Naturalistic Decision Making (NDM) theories of decision making.

Decision making is dependent on an interaction between both the situational demands of the task as well as individual differences in the decision maker. In addition, two key elements of decision making are the accuracy of the decision taken and its associated confidence. The body of work contained in this thesis, therefore, examined the impact of both external factors relating to decision making which included Decision Criticality, Task Load, Time Pressure and Audio communication, as well as the internal factors of Personality, Cognitive Constructs and Video Game experience. These factors were considered in relation to decision accuracy, confidence and W-S C-A (a measure of metacognitive ability).

The experimental stimulus was developed with the help of Subject Matter Experts. This included the development of a realistic computer-generated air defence scenario in which the Task Load (high, moderate, low) was varied. In addition, an appropriate range of classifiable decision options, which varied in Decision Criticality (high, medium, low) was also generated. Three-hundred novice participants and twenty-two experts took part across a range of experiments. The Experimental Work consisted of two sections. The first section contained the foundation and investigatory work. The second section assessed the application of this measure through the use of feedback/training and expert participants.

The results highlight the impact that Decision Criticality, which is the level of consequence associated with a decision, has on individual decision making. Findings show that a decision which was higher in criticality impacted positively on the performance by increasing decision accuracy. Individual participants tended to be less confident in their decisions when responding to decision events of medium criticality. Higher Decision Criticality was also shown to increase metacognitive abilities. That is, individuals were better able to discriminate between their accurate and inaccurate responses. Decision confidence was found to be both relatively stable and high across the experiments, indicating high levels of confidence in decisions, regardless of any other experimental variables including Task Load, Time Pressure and Audio. Such confidence was heightened in individuals who played video games on a regular basis. Internal factors also suggest that there may be individual differences that relate to decision making. Indeed, a higher tolerance of ambiguity may be beneficial in helping individuals deal with uncertain environments, such as in air defence. Although results of personality trait were largely inconclusive, low neuroticism and high conscientiousness appear to be beneficial in critical environments.

Additionally, through the introduction of a metacognitive feedback training element, the research investigated the application of the measure. The results from this study demonstrated that, with metacognitive instruction, individuals displayed improved metacognitive ability. The experimental work also contains the first research to apply this method using active Royal Navy air defence personnel. The results from this experiment replicated the findings of criticality in the novice participants. The results thereby identify how this approach could be applied to air defence settings, and, illustrate the increased ecological validity of the findings.

Recommendations suggest the findings can be applied to training and decision support technology. Further, the outcomes generally support the potential use of more traditional experimental methods alongside naturalistic approaches in critical environments, and that such an approach is warranted. Researchers and practitioners need to consider new approaches to research design to examine decision making in critical environments going forward.

Overall, this thesis further contributes to the air defence decision making domain by providing valuable insights into the external and internal factors that are significant and relate to air defence decision making. Importantly, the work clearly showed Decision Criticality as an important factor that needs to be addressed when investigating decision making in critical environments. Individual differences were also demonstrated to be an important consideration.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Dr Michael Jump and Dr Jacqueline Wheatcroft for their support and guidance. It has been a challenging journey but we made it! Thank you Dr Jump for your help to develop the code required for the experimental scenarios. I would also like to thank Professor Susan Higham for her support in the last few months.

Secondly, this thesis would not be possible without the funding provided by the Defence Science and Technology Laboratory (Dstl). Thank you to my technical partners at Dstl, Robert Walden and Martin Lord. I would also like to thank the Subject Matter Experts at Systems Engineering and Assessments (SEA Ltd.), Jonnie Barnes-Yallowley, Jim Langrishe and Michael Jackson. Thank you for your expert knowledge and the valuable time you gave to help develop the experimental setup.

Thank you to the UoL Flight Science Research Team; Rebecca Mateer, Christopher Dadswell, Josie Roscoe, Neale Watson, Wajih Memon, Dr Michael Kelly and Dr Sarah Scott who welcomed the ‘non-engineer’ into the engineering team. Thank you for providing me with your invaluable support, friendship, brew/AJ breaks and laughs throughout this time, with a special thank you to Christopher Dadswell and Thomas Fell for their ongoing ‘tech’ support. You will be pleased to know that I have not been on a secret mission analysing you all! I would also like to thank my Psychology researchers, Joanne, Grace, Hayley, Sara and Frankie for all their support and advice.

Thank you to all the participants who took part in my experiments and those who gave up their time to help me with the recruitment.

To my friends and family, thank you for your continuous support. A special thank you to my Mum and Dad, I cannot put into words how much I appreciate your love, support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

Term	Brief Description
Anti-air warfare officer (AWO)	An officer who is concerned with the 'bigger picture' for the Task Group which may mean looking at and/or sensing beyond the ship to a radius of up to 200 Nautical Miles.
Classical Decision Making (CDM)	CDM theory postulates that the decision maker is able to weigh up their options and make a decision on that basis. Laboratory based experimentation.
Cognitive Task Analysis (CTA)	A knowledge elicitation method which is aimed at understanding the cognitive processes involved in expert decision making.
Commanding Officer (CO)	The CO is in ultimate command of the ship but will most likely not 'interfere' with the decisions of the AWO and PWO in terms of the maintenance of situational awareness and battlespace management.
Common Operating Picture (COP)	A compilation of data sources which can be visual, sonar, magnetic and satellite.
Counter Threat Tasks	Task which operate in a reactive mode to the perceived threat that has to be dealt with.
Decision Criticality (DC)	A decision event was defined as; an occasion where a decision needs to be made by an operator.

	A decision event with a higher consequence if made incorrectly held a high DC level.
External Factors	Factors which related to the environment, which included, TL, DC, audio and time pressure.
Internal Factors	Factors that related to the individual, which included, personality, cognitive constructs.
MATLAB	A mathematical coding environment.
Metacognition	An awareness of ones' performance, and the ability and willingness to reflect on ones' thinking processes.
Metacognitve Feedback Training (MFT)	A tool developed which included a PowerPoint briefing to increase understanding of the Confidence-Accuracy relationship, combined with an additional briefing provided during the practice phase.
Microworlds	*Microworlds* are simulated task environments and are generally computer-based behavioural simulations which aim to assess dynamic decision making.
Naturalistic Decision Making (NDM)	NDM aims to understand the way people use their experience to make decisions in field settings.
North Atlantic Treaty Organization (NATO)	An intergovernmental military alliance between 29 North American and European

	countries.
Novice	An inexperienced individual in the domain (i.e. air defence decision making).
Operations Room (Ops room)	The focal point in a ship for the conduct of the real-time external battle.
Principal Warfare Officer (PWO)	An officer who is focussed on the self-defence of the individual Unit and hence might be concerned with any activity, displayed on the radar screens, in close proximity to the ship.
Routine Tasks	Tasks which include battlespace management and picture compilation.
Situational Awareness (SA)	An individuals' awareness of their environment.
Subject Matter Expert (SME)	Individual with specialised knowledge and experience in the domain studied.
Tactical Decision Making Under Stress (TADMUS)	A 7-year research programme which aimed to help mitigate the impact of stress on decision making.
Task Group (TG)	A TG will likely consist of a number of different Units (Ships)
Task Load (TL)	The number of tracks displayed on the radar screen and the frequency and speed at which decision events occur.

Unit	A single ship.
Video Game Player (VGP)	Video game players who reported having played video games. VGP1 = *less than 7 hours per week*, VGP2 = *more than 7 hours per week.*
Virtual Avionics Prototyping Software (VAPS)	VAPS XT is a PC-based software tool which was used to generate dynamic, interactive and real-time graphical Human Machine Interfaces (HMI).
Within- Subjects Confidence-Accuracy (W-S C-A)	W-S C-A method is able to calculate the statistical relationship between the levels of confidence individuals might place in responses given relative to the corresponding decision correctness.
Workload (WL)	A construct related to individual's resource capacity and the impact of the stressor on the individual.

PART ONE: Literature Review and Background

CHAPTER ONE:

1. The Problem of Air Defence Decision Making

1.1. Introduction

This Chapter introduces the operational context of air defence by providing a brief background into air defence decision making. The Chapter also sets out the problem statement and describes how this has been addressed through the aims and objectives of the research project. In addition, as this thesis examined the Decision Accuracy, Confidence, and Metacognitive abilities of air defence operators, these terms have been introduced together with their associated definitions.

1.2. Air Defence Definitions and Operational Context

Air defence has been defined by North Atlantic Treaty Organization (NATO) as “all measures designed to mollify or reduce the effectiveness of hostile air action” (AAP-6, p.4). In the air defence role, different levels of decision making are required; these range from a task group (TG) level to a single ship (Unit) level decision making. A TG will likely consist of a number of different Units and thus the air defence task can be considered to be conducted at a number of levels; Force Level (i.e. defence of the TG) and Unit Level (i.e. defence of individual vessels). Each ship may, therefore, have authority for certain tasks delegated to it by the Force Commander. For example, the Type 45 is a specialist air defence vessel and would often be the ship with delegated authority for Force Level air defence (Hunt, personal communication, 2014). In a ship’s operations (Ops) room of a Type 45, Ops room personnel (operators) must manage all information from sensors and data sources. The data sources can be visual, sonar, magnetic and satellite - therefore, fusion and sharing of data is of crucial importance. These data sources are compiled into a Common Operating Picture (COP), which is a single display that provides all of the relevant information to assist the operators in their decision making. This data is displayed to them on a radar screen. As such, a ship's Ops room is the focal point for the conduct of the real-time external battle.

The tasks undertaken by personnel in the Ops room can be categorised into 'Routine' tasks and 'Counter Threat' tasks. Routine tasks include battlespace management and picture compilation. Here, personnel follow Standard Operating Procedures (SOPs) and instructions issued in Operational Tasking Orders. In comparison, for Counter Threat tasks, personnel are operating in a reactive mode to the perceived threat that has to be dealt with. This response can be at both Force and Unit Level. Each ship is commanded by the Commanding Officer (CO) but is supported by the Principal Warfare Officer (PWO) and the Air Warfare Officer (AWO). The PWO is focussed on the self-defence of the individual Unit and hence might be concerned with any activity, displayed on the radar screens, in close proximity to the ship. The AWO is more concerned with the 'bigger picture' for the TG which may mean looking at and/or sensing beyond the ship to a radius of up to 200 Nautical Miles. The CO is in ultimate command of the ship but will most likely not 'interfere' with the decisions of the AWO and PWO in terms of the maintenance of situational awareness and battlespace management. The PWO is responsible for providing advice to the CO by collating all available information to provide a summary of the tactical and general operational situation. Furthermore, within limits of authority, PWOs are responsible for the correct reaction to the operational situation by own ship and own weapons. The scope of this thesis focuses on Unit Level air defence decision making, in Counter Threat tasks, investigating the PWO as the main decision maker.

1.3. Air Defence: Decision Making

Incidents in air defence history have demonstrated how errors in decision making can have disastrous consequences. For instance, in 1982, in the UK, a British Army Gazelle was mistaken for an Argentinian aircraft by Her Majesty's Ship (HMS) Cardiff which resulted in fatal friendly fire contact (Ministry of Defence [MoD], 1986). Another well-documented incident occurred in 1988 when Aegis Cruiser United States Ship (USS) Vincennes shot down an Iranian commercial airliner mistaking it for a hostile aircraft killing 290 people (Fogarty, 1988). The reported timeline of events shows that, initially, the airliner appeared on the USS cruiser's radar as an unknown assumed enemy. This was coupled with confusion as the Identification Friendly or Foe (IFF) displayed the aircraft as both a commercial and military aircraft. In response to this, the USS cruiser sent a warning to the aircraft but failed to receive any response. In addition, there was confusion surrounding whether the aircraft was actually ascending or descending. Final warnings were sent but again no response was received. As a result, the USS Vincennes launched its missiles from 8 NM away at the aircraft which was at an altitude of 13,500ft. Crucially, these events all occurred within the very short time period of 7 minutes (Craig, Morales & Oliver, 2004). The Fogarty report (1988) into the incident assigned human error as the main contributing factor. This general term included poor decision making and erroneous expectancies which were likely to be induced by the stress imposed on personnel. Stress was argued to have arisen from a range of factors such as time-pressure, uncertainty and lack of information.

Such incidents emphasise the need to gain a better understanding of how decisions are made under time pressure and in the dynamic uncertain conditions that surround air defence situations. Subsequently, as a response to the Vincennes incident described above, it was believed that further research was needed to investigate the impact of stress on decision making in critical environments and The Tactical Decision Making Under Stress (TADMUS) research programme was created (Cannon-Bowers & Salas, 1998). TADMUS was a 7-year research programme which aimed to help mitigate the impact of stress on decision making. There were six tasks TADMUS aimed to investigate: 1) definition and measurement of critical decision tasks; 2) examination of stress effects on decision making, 3) development of decision support principles and experimental prototypes, 4) development of training and simulation principles, 5) development of display principles and integration of training and 6) decision support strategies. Some of the main findings to emerge from these efforts include the fundamental and theoretical framework for studying complex, real-life decision making. For many, this was the start of the Naturalistic Decision Making (NDM) paradigm (Klein, Orasanu, Calderwood & Zsambok, 1993) which is discussed in Chapter two. Hence, whilst air defence decision making has been explored in various ways (see Chapter 3), this thesis aimed to specifically look at decision accuracy, confidence and metacognitive abilities of air defence operators and to work toward the development of a toolkit to assess some elements of these aspects which have not previously been conducted.

1.4. The Problem Statement

Trained air defence operators must detect, locate and identify potential air threats, coupled with complex and cognitively demanding decisions in the dynamic and uncertain environment of naval warfare. This environment often involves information overload, time pressure and ambiguous information. Additionally, in air defence decision making, there are many different factors that play a potentially significant role in an operator's decision making ability. These factors can include both external factors from the environment, such as Task Load and internal factors such as, individual differences, which interact to inform decision making. Although there has been substantial work in the area of decision making in critical environments, the current thesis has examined the interplay of both the environmental and individual interplay in decision making and how metacognitive ability might be measured and assessed in these environments.

Two key elements of decision making are the accuracy of the decision taken and its associated confidence. Further, no research has examined levels of confidence individuals have in their own decisions with specific regard to complex environments such as onboard a ship using a metacognitive measure that has been used reliably in other contexts to assess relationships between confidence and accuracy (Wheatcroft, Jump, Breckell & Adams-White, 2017).

1.5. Decision Confidence and Decision Accuracy

The thesis has specifically looked at individual accuracy and confidence in decisions in air defence with corresponding metacognitive ability. Decision confidence in one's own ability plays an important role in the decision made (Griffin & Tversky, 1992) and assessments of confidence can be used to guide current and future decisions (Kepecs & Mainen, 2012). Research into decision confidence derives from perceptual choice tasks, i.e., Signal Detection Theory (SDT – see Chapter 3 for discussion). Recent research conducted in perceptual choice tasks found that people seek further information when they lack confidence in their initial choice (Desender, Boldt, Yeung, 2018). It is therefore important that air defence operators are able to apply the correct amount of confidence to a decision. For instance, overconfidence is related to risk-taking behaviour (Lovallo & Kahneman, 2003). On the other hand, under confidence tends to lead people to seek further information to make their decisions (Lanzetta, 1963). Confidence may be influenced both by the environment i.e. external as well as the individual i.e. internal.

With regards to external factors, research has shown that the amount of information available to the decision maker (Hammer & Ringel, 1965), the certainty of that information (Sieber & Lanzetta, 1964), the level of decision danger (Wheatcroft et al., 2017) and difficulty of the decision, i.e. the hard-easy effect (see Wheatcroft, Wagstaff & Manarin, 2015) impacts on decision confidence. The air defence environment is uncertain, dynamic and ambiguous; thus increased understanding of how these external factors impact on an individual's ability to make confident decisions can influence effective decision making. Confidence has also been attributed to individual factors. Kleitman and Stankov (2001) describe this as a confidence factor. For example, it has been argued that there is a confidence trait which remains stable across a range of difficulty levels and by definition is not state-dependent (Pallier et al., 2002). Further, individuals vary in their ability to estimate the reliability of their own decisions (Song et al., 2011). Research has also found confidence to be related to other concepts such as cognitive ability and personality (Kleitman & Stankov, 2007; Pallier et al., 2002). Moreover, confidence as an individual trait has led some researchers to argue that confidence can be a good behavioural predictor. In other words, there is a predictive validity of confidence bias which can be generalizable across domains (Pallier et al., 2002). In keeping with this, decision confidence may be a useful tool for selection and training purposes in the air defence domain.

However, one of the most consistent and robust findings is that people tend to overestimate their ability (Tversky & Kahneman, 1974; Wheatcroft, Wagstaff & Kebbell, 2004). This leads individuals to believe that they know more than they do. It is, therefore, crucial to examine the corresponding accuracy of that decision. This is known as the confidence-accuracy (C-A) relationship. C-A has been previously investigated in different domains which include, for example, eyewitness testimony (Wheatcroft et al., 2004), and education (Nietfeld, Cao & Osborne, 2005). A good positive relationship between confidence and accuracy is highly beneficial as it suggests that individuals weight information and decisions appropriately and place them in perspective with respect to data from other sources (Stichman, 1967). Individuals thereby apply the correct amount of resources to a decision in a time-effective manner without displaying an over / or under confidence, however, little is known about the C-A relationship in critical environments such as air defence decision making. One way that one can conceptualise the C-A relationship is as a metacognitive activity.

1.6. Metacognition

The term metacognition refers to an awareness of ones' performance, and the ability and willingness to reflect on ones' thinking processes (Parker & Stone, 2014). It has been argued, therefore, that metacognitive judgement is an important psychological construct that should be included in the study of decision making processes (Jackson & Kleitman, 2014). Additionally, Fleming and Lau (2014) state that, the relationship between decision confidence and accuracy can be used to provide a quantitative measure of metacognition.

Metacognition is multifaceted, and for many, it is considered to have two elements, (i) metacognitive regulation and (ii) metacognitive awareness (Flavell, 1979). Specifically, metacognitive regulation refers to an individual's ability to monitor their knowledge and control cognitive processes (Flavell, 1979). Thus, one aspect of regulation refers to the experience of the feeling of confidence which is part of a larger aspect of cognitive self-monitoring (Efklides, 2001). Hence, metacognition can be assessed using decision confidence. Although there have been many ways of measuring metacognition, which are discussed in Chapter three, this thesis assesses the C-A relationship in decision making by using a measure of Within-Subjects Confidence-Accuracy hereinafter, referred to as W-S C-A. The W-S C-A is calculated using a measure of confidence and the corresponding accuracy of the decision made. This measure has been successfully used in other domains (Wheatcroft et al., 2017; Wheatcroft & Woods, 2010; Wheatcroft et al., 2004).

In a similar fashion to decision accuracy and confidence, metacognitive abilities can be influenced by a range of external and internal factors. The external factors include, for example, experience (Wheatcroft et al., 2017) and cognitive load (Jackson, Kleitman & Aidman, 2014). Importantly, individual differences in metacognitive abilities in air defence operators have not been investigated in the air defence domain. It has been argued that metacognitive ability is independent and that a metacognitive trait mediates the accuracy of self-assessment. Hence, there is a consistent confidence level irrespective of accuracy (Pallier et al., 2002). Further to this, it has also been argued that metacognitive ability is a relatively stable construct of personality that can be quantified and made subject to training and improvement (Jøsok et al., 2016). Hence, by using the W-S C-A measure this research aimed to increase understanding of the air defence domain by the introduction of a novel approach to performance measurement in air defence decision making.

Consequently, the fundamental problem that this thesis addresses is to increase the understanding of, and measure, decision confidence and accuracy in air defence decision making. This was achieved by addressing the following four broad aims:

1. To investigate both internal and external factors that impact on decision accuracy and confidence within the context of air defence.

2. To expand on previous measures of performance and confidence by examining and measuring the metacognitive ability of individuals in a critical environment through the application of a novel measure.

3. To assess the application of this measure through the use of training/feedback and expert participants.

4. To take first steps to the development of a Toolkit to measure decision accuracy, confidence and metacognitive abilities of air defence operators.

1.7. Orientation and Outline of the Thesis

The current thesis examines decision accuracy, decision confidence and the relationship between accuracy and confidence as it pertains to metacognitive ability in an air defence domain; namely the ship's Ops room. To do so, the body of work contained in the current thesis aimed to establish some of the potential factors which are related to accuracy, confidence and metacognition, in addition to the development and design of a novel experimental paradigm. The thesis, therefore, has introduced a method to examine decision making in the air defence environment to examine the metacognitive abilities of operators through the use of the measurement of W-S C-A. The measure has been successfully applied in other critical domains. It is envisaged that the outcomes of the research contained within the current thesis may be used to help prioritise selection, training, and identify individual needs in order to improve the effectiveness of decision making in naval air defence sceanrios and determine how decision support tools might best be used.

To address the aims outlined above, the thesis is presented in three parts. Part One contains the introduction, relevant background and the development of the methodological approach. In Part One, Chapter two aims to establish the factors that may relate to decision accuracy and decision confidence and metacognitive ability. Chapter three reviews the methods previously used to assess decision making and measures of confidence and accuracy. Chapter three also provides evidence of the validation of the measure by the way of a previously published paper. In Chapter four, the method and process of developing the measure and stimulus development is described. This methods Chapter concludes the first part of the thesis by describing the materials, design, participants, and procedures employed in the Experimental Work. The second Part contains the Experimental Work, which addresses the aims of the thesis and the third part of the thesis is comprised of the application of the findings together with limitations of the work, implications and recommendations based on the research outcomes. The following Section will provide an outline of the thesis in more detail.

PART ONE: Background and Literature Review

Chapter One provides an introduction to the context of the current thesis. Air defence decision making is complex, dynamic and time pressured. Hence, understanding the interplay of factors that influence decision making is important to ensure effective decisions are made in this critical environment. This first Chapter provides an overview of the operational air defence context, the problem statement, the scope of the thesis and defines key terminology.

Chapter Two introduces the theoretical backdrop to the research. The current thesis has taken an integrative approach and examined decision making in the air domain context from perspectives of both Naturalistic Decision Making (NDM) and Classical Decision Making (CDM) theories. In addition, decision making often involves the interplay of different factors and the second Chapter discusses the potential external and internal factors to help establish how these factors might relate to air defence decision making. The current research investigated the impact of Decision Criticality (DC) Task Load (TL), Time pressure, Audio in addition to Personality, Cognitive Constructs and Video Game Play.

The current thesis has introduced a new method to measure accuracy, confidence and metacognition in air defence personnel through the combination of elements of experimental laboratory testing with NDM methods. Hence, **Chapter Three** critically examines the previous methods used to investigate decision making in critical environments and thereby inform the current work.

Chapter Four details the development of the experimental scenario. To ensure high ecological validity a significant amount of effort was spent on developing the materials and scenarios. The development was broken down into five phases and finalised with a pilot study. The method was validated by a previously published paper by the author. The fourth Chapter concludes Part One of the current thesis. Part Two of the thesis contains the Experimental Work which is divided into two sections. Section A: Foundation and Investigatory Work (Chapters 5-8) and Section B: Application (Chapters 9-10).

PART TWO: Experimental Work

SECTION A: Foundation and Investigatory Work

Chapter Five contains the first experiment, which investigated operator's decision making during a computer-based air defence scenario. In order to assess the impact of comparative groups, DC and TL on decision making, outcome measures of accuracy, confidence, and W-S C-A were calculated. In addition, personality, cognitive constructs, workload (WL), situation awareness (SA) were also recorded to assess relationships between the factors and confidence as they, in particular, relate to accurate decision making in air defence. To further increase understanding of both individual and external factors related to Ops room decision making, this first study compared a student sample of Non-Gamers (NGs), with Video Game Players 1 (VGP1) and Video Game Players 2 (VGP2). VGP1 consisted of video gamers who played on average more than seven hours per week and have done so for the previous two years, VGP2 consisted of video gamers who played less than seven hours per week. Findings from the first experiment suggested that DC played a key role in decision accuracy and confidence. W-S C-A was found to be relatively poor and uninfluenced by any factors. Despite this, some individual differences in groups and cognitive constructs were observed.

Chapter Six reports on Experiment 2 which followed from Experiment 1 to examine the impact of time pressure. The same variables were used as described in the first experiment. However, the time to make a decision was reduced from 20 seconds to 10 seconds. The results mirrored those found in the first experiment in terms of the impact Decision Criticality (DC) had on decision accuracy and confidence. To further examine the impact of time pressure, **Chapter Seven** analysed the differences between the first experiment and the second experiment. The analysis found that the reduction in time did not influence decision accuracy, confidence or W-S C-A. However, differences were now observed in the subscale scores SA. The findings suggest that the time reduction did increase operators' feelings of attentional supply and demand.

Chapter Eight considered a further factor relevant to the environment and reports Experiment 3 which investigated the introduction of an audio stimulus on decision accuracy, confidence and W-S C-A. In doing so, participants were instructed to either attend to the audio or instructed that it was background noise. Although the task manipulation of audio was unsuccessful, differences in WL were found in individuals who were instructed to attend the audio. This third experiment demonstrated additional support for the impact of DC on decision accuracy and confidence found in the first two experiments.

SECTION B: **Application**

The aim of Experiment 4, and which is reported in **Chapter Nine,** was to investigate the application of the findings from the first three experiments and form the basis of Section B of the Experimental Work. Part of this section introduced a Metacognitive Feedback Training (MFT) element to the task to examine whether this kind of element impacts on participants W-S C-A (e.g., metacognitive sensitivity). An additional group of participants received MFT prior to the same task employed in the first experiment. The outcomes of the group who received MFT were then compared to a randomly selected sample from Experiment 1. The findings of this experiment showed a positive impact of MFT by a significant increase in the W-S C-A relationship in the MFT group. Importantly, MFT increased confidence in correct decisions. However, all participants displayed high levels of confidence in their decisions.

To increase understanding of the application of the outcomes of the previous Experimental Work contained in the thesis, and to improve the validity of the findings**, Chapter Ten** reports on the results of Experiment 5. In addition to gathering novice data from the previous experiments, a sample of military data, whilst very difficult to obtain, was achieved. The fifth experiment investigated Royal Naval personnel who had recently completed a PWO training course. Although this experiment was conducted on a small sample of experts (i.e., twenty-two PWOs), results showed support for the previous findings in novice participants, increasing the validity of those outcomes.

PART THREE: General Discussion, Limitations, Implications, Recommendations and Conclusions

The third part of the thesis contains a summary of the findings together with a general discussion, limitations and implications of the findings. **Chapter Eleven** discusses the aims of the current thesis relative to those findings and how external and internal factors impact on decision accuracy and confidence in air defence decision making outcomes. Importantly, it emphasises the need for future research to examine the key factor of Decision Criticality. Broadly, the results consistently demonstrate that decisions made in low criticality classifications have the potential to induce feelings of overconfidence. On the other hand, decisions classified as medium criticality provide for uncertainty showing a reduction in confidence, but not accuracy. Metacognition, as measured by W-S C-A, was shown to be relatively low throughout the experiments with a general tendency for overconfidence. Nevertheless, the outcomes from the fourth experiment illustrate that W-S C-A can be improved with a specific MTF session prior to the start of the task. Furthermore, individuals' confidence in inaccurate decisions remained high regardless of any external factors which suggest that decision confidence is a stable trait. Individual differences in decision making varied across the experiments. One of the main and consistent findings was that individuals with a higher tolerance to ambiguity performed more effectively in the tasks.

Chapter Twelve discusses the limitations, implications, recommendations, future work and conclusions. The methodological approach in the current thesis integrated both NDM and CDM theories. Drawing on a range of current thinking to inform the methodological stance has demonstrated that an integrative approach is useful to the decision making research domain and provides valuable insights to the benefit of both approaches. This Chapter discusses the implications for research which might include, for example, assistance for decision supports that may highlight and identify the dangers of overconfidence in certain situations. Confidence was found to be unrelated to TL and, as such, could arguably remain a stable trait over the course of a scenario and/or critical incident. The work contained in this thesis points toward further research into MTF to improve trainee and operator metacognitive insight to ensure effective decisions.

The last Chapter discusses the research contained in the thesis which provides valuable insights into the factors that influence decision accuracy, confidence and W-S C-A metacognition. Importantly, the work highlights the consistent impact that DC has on outcomes in air defence and which have not been investigated previously. Overall, the findings would be of benefit to decision supports and enhance aspects of the selection and training of personnel in air defence. Further, the finding that MFT increased metacognitive awareness is beneficial. As such, it should be included in any future training protocols in the air defence context to help increase confidence in accurate decisions and minimise overconfidence in inaccurate decisions taken. Finally, individual traits should not be overlooked.

CHAPTER TWO

2. Theories of Decision Making and the Factors that Influence Decision Making in Air Defence

2.1. Introduction

The purpose of this Chapter is to build a foundation for the theoretical aspect to the thesis and the theoretical approaches relevant to the context of air defence decision making. As such, this Chapter will first discuss two theories of decision making which are relevant to this thesis. These theories also provide the backdrop for the methodology used in the research (as discussed in Chapter 3). Two of the main and competing theories of decision making are provided by Classical Decision Making (CDM) and Naturalistic Decision Making (NDM; Klein et al., 1993). The research adopted an integrative approach to move toward the development of a measure to investigate decision making in air defence. As reported in Chapter one, one aim of the thesis was to establish the factors related to decision accuracy and decision confidence together with metacognitive ability as relevant to air defence decision making. This Chapter reviews the current literature surrounding potential external and internal factors involved in the decision making in the air defence domain.

2.2. Theories of Decision Making

2.2.1. Classical Decision Making: CDM theory postulates that the decision maker is able to weigh up their options and make a decision on that basis. According to classical theories the decision maker selects the optimal decision option from a set of alternative decision options. In this instance, decision makers are considered rational. Based on the concept that decision makers are rational, some authors, therefore, argue that generally, people aim to make accurate decisions with optimal outcomes (Hammond, 2000). This further implies that individuals also aim to make optimal decisions in optimal environments. However, decisions tend to be bounded in their rationality and hence not optimal in nature (Simon, 1955). Bounded rationality is the concept that the decision maker cannot make decisions without some form of cognitive limitations (Simon, 1955). Such a concept can be further expanded to include ecological rationality which explains how decisions are made that are saturated with contextual and situational pressures, again reducing optimal decision making (Todd & Gigerenzer, 2007). As described in Chapter one, the air defence environment is uncertain and thereby optimal decision making is made more difficult. Lipshitz and Strauss (1997, p.150) define uncertainty as a "sense of doubt that blocks or delays action". Lipshitz and Strauss (1997) also identified three types of uncertainty: inadequate understanding, incomplete information and undifferentiated alternatives (i.e., those too close to each other to determine an optimal choice). Henceforth, to facilitate decision making in real-world situations, people simplify complex judgements by employing shortcuts to reduce cognitive effort through the use of cognitive heuristics (Kebbell, Muller & Martin, 2010). Cognitive heuristics are rules of thumb which assist in processing information and include confirmation bias, representative bias, availability, representative and anchoring (Tversky & Kahneman, 1974). Fast and frugal heuristics build upon heuristics research by investigating when, how and why heuristics assist people to make decisions and includes the concept of ecological rationality to its research (Gigerenzer & Todd, 1999). Thus, as part of the decision making process, individuals use these cognitive processes and resources to infer meaning from the information gathered.

Consequently, the classical approach to decision making tends to prescribe how individuals should make decisions with a focus on the outcome of the decision. However, Beach and Lipshitz (2017) conclude that the classical approach to decision making is inappropriate in real-world environments, such as air defence. With regards to the research methodology, classical approaches are generally laboratory-based and may also include the production of analytical models (Reimer & Rieskamp, 2007). The population the research is conducted on is generally student and non-experts, thus allowing for large sample sizes. For these reasons, classical approaches are high in internal validity in that they are replicable, and involve controlled environments with variable manipulation. However, this methodology is limited by its external validity.

2.2.2. Naturalistic Decision Making: To counter some of the limitations of the classical approach such as the use of non-experts, forced choice and laboratory-based research designs, Klein and colleagues (Klein et al., 1993) introduced NDM. NDM defines decision making as "a commitment to a course of action that is intended to yield results that are satisfying for a specified individual" (Yates & Tschirhart, 2006; p.422). As described in Chapter one, NDM aims to understand the way people use their experience to make decisions in field settings (Zsambok & Klein, 1997). As such, NDM research endeavours to be high in ecological validity and is domain specific. NDM investigates how experts make decisions in environments that have been defined as ill-structured, uncertain, ill-defined, high stakes, include feedback loops, organizational goals and norms, and time-stressed (Gore, Flin, Stanton & Wong, 2015; Orasanu & Connolly, 1993). Furthermore, NDM attempts to understand human capabilities and is interested in decision making processes, not just the outcomes.

Consequently, NDM models are descriptive and aim to uncover what information decision makers use, how they interpret it and which decision rules they may apply. NDM differs from the more classical approach to the investigation of decision making as it relies on experts, field settings and self-reported measures. As such, classical theories do not accurately describe what people actually do to make decisions. Several theories have also been developed from NDM. These include Recognitional Primed Decision making (RPD). RPD describes how people use their experience in the form of a repertoire of patterns (Klein, Calderwood & Clinton-Cicorro, 1986). Accordingly, decision makers highlight the most relevant cues, provide expectancies, identify plausible goals and suggest typical types of reactions in the type of situation present which enables them to match the situation to the pattern they have learned. If it is a clear match they then carry out the most typical form of action. Further, experts use mental simulation to imagine how a scenario would progress within the context of the current situation based on a blend of intuition and analysis (Klein et al., 1986). Consequently, NDM provides a sound basis for understanding decision making in air defence, evident by the body of research conducted by TADMUS and in other high stakes environments. However, NDM has been criticised for its limited accessibility to experts, small sample sizes and reduced internal validity (Markman, 2018).

2.2.3. Integrative Approach: The CDM and NDM approaches described both aim to increase the understanding of how decisions are made in their own right. However, instead of taking the view that these two approaches sit separately, more recently, some authors have argued that the two approaches have shared concepts and could, therefore, benefit from integration (Klein, 2015; Roberts & Cole, 2018; Markman, 2018; Bartels, Hastie & Urminsky, 2018). Subsequently, instead of approaching research from one or the other paradigm, this thesis seeks to integrate some of the NDM concepts with classical methodologies to examine and measure decision accuracy, confidence and metacognition. It is envisaged this will provide a wider understanding of decision making and metacognitive abilities of air defence operators. Indeed, Lipshitz, Klein, Orasanu and Salas (2001) argued that NDM should include both laboratory settings as well as the more naturalistic observations. Furthermore, it has been argued that, to help develop NDM theory, more experimental work could be used to help make predictions (Orasanu et al., 1998) which can be used alongside more traditional NDM methods. It has been argued that research should be guided by the need to drive and improve the credibility and transferability of methods used in NDM and, importantly, these approaches should be used in conjunction with NDM, methods not in opposition to (Roberts & Cole, 2018; McAndrews & Gore, 2013).

Chapter three will discuss further the benefits of using an integrative approach and introduce the means by which to measure decision accuracy, confidence and metacognitive abilities of operators. Thus, the work contained in this thesis integrates aspects of CDM laboratory studies such as, behavioural measures, narrow focus (manipulated variables) with some NDM criteria including; self-reports and complex decision environment (see Table 1). In addition, a mixture of both approaches was incorporated into the design to increase both internal and external validity, as well as the use of both novice and expert participants.

Table 1: Comparison of the characteristics of Laboratory and NDM traditions (taken from Markman, 2018).

Laboratory Studies	**NDM**
Focussed narrowly	Focussed broadly
Novice participants	Expert performance
Ignores context	Incorporates context
Focussed on behavioural measures	Focussed on self-report measures
Simple environments	Complex environments
High internal validity	High external validity
Supports theories of decision making components	Supports integrative theories

2.3. Factors Relating to Decision Making in Air Defence

The effectiveness of decision making can be influenced by many factors. In order to understand decision making, and begin to move toward the development of a Toolkit to measure decision accuracy confidence and metacognition, it is important to understand some of the potential factors that may impact on decision making. It has been argued that decision making depends on three categorical factors; 1) task complexity, 2) environmental conditions, and 3) person characteristics (Einhorn, 1970). The concept that decision making does not occur in a vacuum is also expressed by the Social Cognitive Theory (Bandura, 1993) which states that there is a triadic reciprocal causation between the individuals, environment and the outcome of behaviour. As such, it seems natural and evident that both internal factors from the individual and external factors from the environment interact to inform decision making. In light of this, this Chapter discusses the factors that potentially influence air defence decision making in relation to both external and internal factors.

2.3.1. External Factors: The external factors that surround air defence are complex due to the unpredictability and the continuously changing environment of a ship's Ops room. It is therefore important for the operator to adapt to these changes whilst maintaining effective decision making. For example, in air defence, external factors may involve the rapid change in speed and altitude of an aircraft, and the number of aircrafts on the radar screen which must be attended to. Additionally, external factors can also involve the task situation such as a peace enforcement or wartime mission, group dynamics, amount of information and information ambiguity. A number of factors were investigated, these included Task Load (TL), Decision Criticality (DC), time pressure, and audio communications, to examine the extent to which these factors may influence decision making in respect of decision accuracy, confidence and metacognition.

2.3.1.1. Decision Criticality: An external factor which may be relevant to air defence decision making is decision criticality (DC). Hanson, Bliss, Harden and Papelis (2014) stated that criticality is important to decision making, but has been ill-defined in the literature. In this thesis DC has been defined as the associated consequence of an individual decision i.e. a decision event, which varies in levels to include low, medium and high DC events. For example, a decision event with a higher consequence if made incorrectly held a high DC level (see Chapter 4 for a more detailed description). In an air defence scenario, criticality plays an important role and operators need to be accurate and confident in the decisions made. As previously mentioned in Chapter one, air defence decision making involves a range of tasks. Some of these decisions will be low in criticality, i.e. Routine tasks. In comparison, Threat Response tasks, have a much higher criticality associated to it. Fundamentally, therefore, there is dearth of research on the impact of the criticality of different types of decisions in air defence. To the author's knowledge, only one study has looked at the relationship between decision criticality and decision accuracy, confidence and metacognition (see Adams-White et al., 2018; this paper is a part-product of this thesis). Hence, this thesis builds on the knowledge of how criticality impacts on decision making in critical environments.

Nevertheless, there has been some literature which has examined decision making and criticality with regards to responses to alarm systems. For instance, Bliss and McAbee (1995) examined whether the criticality impacted on response to alarms. In this study, participants were required to take part in a spatial orientation task as well as respond to alarm systems. To respond to the alarm they were required them to move the cursor arrow and click the mouse within 15 seconds. Criticality was induced by informing participants that more points would be deducted from their overall performance score in the highly critical condition. Although the manipulation of criticality was not successful (i.e., the findings were not significant), the results did show that, in low alarm criticality, participants made more accurate responses to the alarm. In another study, Hanson et al., (2014) successfully manipulated criticality by instructing the participants prior to the task. High criticality was defined as "a task that has potentially life threatening consequences" (Hanson et al., 2014, p. 2). In this task high, criticality also included more time pressure. The study found that participants made more errors in the highly critical scenario. It was therefore concluded that high criticality has a negative impact on decision making, supporting the findings of Bliss and McAbee (1995). Nevertheless, research into criticality has also shown that critical tasks do not impair cognitive ability (Callister, Pervival & Retzlaff, 1999). Consequently, criticality may also have a positive impact on performance.

The previous studies examined the criticality of the scenario and did not specifically investigate the criticality of an individual decision event. The individual decision events are also important to consider when investigating decision making. Indeed, previous work has considered decision danger, which is the level of risk associated with a decision, and decision difficulty. Both the difficulty of decision and decision danger (Wheatcroft, Wagstaff & Manarin, 2015; Wheatcroft et al., 2017) have been shown to impact on decision confidence and accuracy. The aim, therefore, is to begin to build an understanding as to when errors in decision accuracy and confidence may occur. As such, this thesis has specifically looked at the criticality of each type of decision, not just of the scenario. Furthermore, in the previous studies, participants were informed about the criticality of the scenario. The research was interested in the perceived criticality of the decisions as operators may not necessarily be aware of criticality at the time of making a decision.

2.3.1.2. Task Load (TL): Understanding the influence of TL is crucial. In a ship's Ops room, TL is variable and operators must make effective decisions in varying TL conditions. Deck and Jahedi (2015) found that individuals made poorer decisions under conditions of high cognitive load. TL is defined in this thesis as a combination of the number of tracks displayed on the radar screen. As mentioned in Chapter one, the radar screen is the focal point of the ship which includes information to assist in their decision making. The tracks on a screen correspond to the experimental manipulation of either a data link, which is the presentation of information received from the data sources of the ship on the radar screen, or an aircraft displayed on the radar screen, also referred to as a Track. This in turn influences the amount of information that the individual must attend to, and the frequency of the decisions during the scenario. Subsequently, a high TL is associated with an increase in temporal demand and cognitive load for the operator which influences the workload (WL) of the individual (WL is discussed in greater detail in Chapter 4). This idea aligns with research into air traffic control (ATC), as it is common that the number of aircraft used is a variable to manipulate TL, which then has an impact on the WL of the task at hand (Friedrichm, Biermann, Gontar, Biella & Bengler, 2018). What's more, an important distinction has been made between TL and WL. TL is the external stress or demands, i.e., the complexity of the task, time pressure, and so on. In comparison, WL is the impact of the stressor on the person however; it is also sensitive to variations in TL (Selcon, Taylor & Koritsas, 1991).

Previous research which has examined TL in air defence defined TL as the number of tracks presented on the screen (3 = *Low*, 6 = *High*; Loft, Sadler, Braithwaite & Huf, 2015). These researchers found that high TL impacted on performance by reducing response time as individuals were slower to make decisions. In addition, the authors also found that TL was increased feelings of subjective WL. Hence, it is not surprising that high cognitive load associated with high TL has also been found to increase feelings of stress/arousal (Miyake, 2001). Research has shown that cognitive ability may be impaired in conditions of high arousal (Darke, 1988; Arnsten, 2009). Indeed, with regards to decision making, heightened anxiety associated with higher arousal has been shown to reduce decision accuracy (Cumming & Harris, 2001; Keinan, Friedland & Porath, 1987; Klein, 1996). Hence, TL can impact on performance by influencing levels of arousal in the individual.

In relation to decision confidence, stress has been shown to increase confidence in decisions (Schaeffer, 1989; Heereman & Walla, 2011). Importantly, if this is not related to accuracy, this can lead to an overconfident assessment of the individual's own ability. Further, emotional states such as stress/arousal have been shown to impact cognitive evaluations of risk (Johnson & Tversky, 1983). As a consequence, stress and arousal have been shown to lead to more risk-taking behaviour (Heereman & Walla, 2011) which may, in turn, lead to greater levels of confidence but not the associated accurate decisions. However, there is also research which has linked cognitive load to more risk-adverse behaviour (Deck & Jahedi, 2015). One possible explanation for the impact of arousal on performance is the Yerkes Dodson law (1908). This theory argues that, when arousal is too high or too low, performance is negatively impacted and that there is an optimal arousal level taking the shape of an inverted 'U'. Moreover, in a more recent study investigating cognitive load, it was found that periods of low cognitive load had a negative impact on performance (Jackson et al., 2014). The research on TL suggests implications for the impact of TL on decision accuracy and confidence.

There has also been some research which examined the conditions in which confident and accurate decisions are made. Jackson et al. (2014) found individuals improved decision accuracy when in conditions of both moderate cognitive load and motion of a simulator (in this study, motion was a task manipulation to induce arousal). This study further examined metacognitive ability and found that low cognitive load negatively impacted on metacognitive monitoring. Crucially, it was demonstrated that individuals were unable to detect changes in their accuracy under periods of low cognitive load or arousal. Hence, low cognitive load was found to be detrimental to metacognitive ability. However, recent research by Adams-White et al. (2018) demonstrated TL did not impact on decision confidence, accuracy or metacognition. An explanation could be provided by the fact that the TL manipulations were not found to be successful. Nevertheless, further research is warranted into the conditions which mediate the relationship between confidence and accuracy and the impact of TL on decision making in the air defence context.

2.3.1.3. Time Pressure: The Ops room is a time-critical environment and decisions must be made quickly, confidently and accurately. For example, the entire Vincennes incident (as described in Chapter 1), only lasted around 7 minutes and this time factor played a role in the decision making process (Craig, Morales & Oliver, 2004). Indeed, time constraints have been shown to impact on how individuals make decisions. This has been demonstrated by individuals employing different decision making strategies in response to the time constraints of a task (Hu, Wang, Pang, Xu & Guo, 2015; Gonzalez, 2004). For example, research has demonstrated that, with greater time constraints, individuals use more simple heuristics (i.e., evidence-based rules) to assist them in the decision making process (Gonzalez, 2004; Gigerenzer & Todd, 1999). Furthermore, some research has found a negative effect on individual’s ability to make decisions effectively (Maule & Edland, 1997; Maule & Svenson, 1993). Gonzalez (2004) determined that, in dynamic decision making environments, time constraints have serious detrimental effects on performance. On the contrary, Kerstholt and Pieters (1994) showed that, in certain tasks, cognitive performance actually improved under time pressure and individuals processed and integrated information more quickly and more accurately. However, in doing so, individuals applied more effort to the task. Hence, time pressure may also increase the experienced WL of the task which, as previously discussed, has been linked to increased feelings of stress (Keinan et al., 1987).

With regards to decision accuracy, confidence and metacognition, Petrusic and Baranski (2003) argued that time is important in improving understanding between confidence and accuracy. It was shown that a longer time permitted on a task increases individuals feelings of confidence. Nevertheless, in a high fidelity medical scenario, which aimed to examine a clinician’s relationship between confidence and accuracy, Yang, Thompson and Bland (2012) found no effect of time pressure on accuracy, confidence or metacognition. However, the results from this study did find an interaction in the difficulty of the task and time pressure. In easier tasks, time pressure increased feelings of confidence, however, in more difficult tasks time pressure decreased confidence. Despite these findings, Gonzalez, Vanyukov and Martin (2005) argued there is little research in dynamic task environments, such as air defence, which has investigated the impact of time pressures. As such, research is needed in more dynamic and realistic scenarios to assess the impact of, and relationships between, decision accuracy and confidence. Furthermore, one of the requirements of the naturalistic decision environment, set out by Orasanu & Connolly (1993), is that the decision making environments are time stressed. Hence, to increase the external validity, it would be important to include time constraints on the decision making task. Additionally, in this thesis, time pressure also related to the manipulation of TL. High TL included more time pressure as the frequency of decisions was increased and therefore increasing the temporal demand on individuals.

2.3.1.4. Audio: Decision making in the Ops room is based on a range of different modalities which includes auditory communications. This is particularly prevalent in an Ops room environment, as there is continuous background noise elicited from alerts, radio communications, and so on.

Research has also shown the presence of audio impacts on performance (Reinten, Kort, Hornikx & Kohlrausch, 2017). An explanation for this is that it interferes with cognitive processing. For instance, the presence of irrelevant radio messages negatively impacts on memory performance (Banbury, Fricker, Tremblay & Emery, 2003). Additionally, extraneous speech has an adverse effect on recall and recognition of complex visual information (Marsh et al., 2015). This has clear implications for Ops room decision making as decisions are made based on both visual and auditory information. Moreover, research has been conducted on the impact of audio warnings and the design of decision support tools (Vachon, Tremblay, Nicholls, Jones, 2011). This line of research demonstrated that different types of auditory alarms impact on the operator's decision making ability. Further to this, research findings demonstrated that there were 22% undetected critical changes provided by audio messages and audio messages lead to a bias in threat detection and objects were perceived as more threatening (Chamberlands, Hodgetts, Vallieres, Vachon & Tremblay, 2018). The influence of audio has clear implication to Ops room decision making. However, little is known about the impact of audio on an operator's decision making ability with regards to decision accuracy, confidence and metacognition in more complex environments.

To recall, metacognition is comprised of metacognitive regulation and metacognitive awareness (i.e., monitoring). Audio may impact on an individual's ability to monitor their ability in the task, as metacognition refers to an individual's ability to monitor their cognitive processes. Indeed, some authors argue that the presence of audio may act in a similar way to that of divided attention. Divided attention refers to processing multiple sources of information to carry out a task (Kahneman, 1973) and research into divided attention has demonstrated that task performance is generally impaired (Kahneman, 1973). Furthermore, individuals have been shown to be unaware of the impact of divided attention and the impact on performance through a lack metacognitive insight (Finley, Benjamin, McCarley, 2014). In other words, the presence of audio affects performance and operators may not be aware of the risks (Finley et al., 2014).

Furthermore, auditory distraction has been shown to impair both memory performance and metacognitive ability by reducing confidence in responses, but not necessarily accuracy (Beaman, Hanczakowski & Jones, 2014). This research seeks to increase the understanding of the effects of audio input on decision accuracy, confidence and metacognition by applying it to an increased fidelity scenario in comparison to perceptual laboratory based work. This is particularly relevant as research conducted thus far has generally focussed on laboratory based experiments using memory encoding tests. Little is known of audio effects and metacognition in the context of higher fidelity experiments. Consequently, there are a range of external factors that may impact on air defence decision making. Commonly, these factors are derived from the situation or decision making environment. As discussed, DC, TL, time pressure and audio were taken into consideration by the research of this thesis. However, as previously mentioned decision making is also influenced by the individual. As such, the internal factors of air defence decision making are an important consideration.

2.3.2. Internal Factors: Internal factors relate to the characteristics of the individual. Specifically, individual differences refer to how individuals differ from one another and research has shown that individual differences play a key role in decision making (Jackson & Kleitman, 2014). Much of the current research examining improving air defence decision making focusses more on systems, such as improving decision support systems and interface design (Vachon et al., 2011) but have not specifically examined how operators interact and the personal characteristics of those operators. This is important as it has been argued that individual differences in cognitive abilities impose greater effects than those due to interface design manipulations (Rodes & Gugerty, 2012). Furthermore, in air defence, Roux and Van Vuuren (2007) argue that operators use their own experience to determine threat evaluations making it difficult to describe the threat evaluation process. Therefore improving understanding of individual differences in air defence operators may shed light on the decision making process and how to improve decisions made. Specifically, in relation to decision confidence and decision accuracy, people vary in their ability to estimate the uncertainty and reliability of their choices as well as their ability to estimate the reliability of their own decisions (Fleming, Weil, Nagy, Dolan & Rees, 2010; Song et al., 2011; Wheatcroft et al., 2017). Interestingly, individual differences have been used as a basis for job selection and training. As such, cognitive ability has been linked to air traffic controller performance and individual differences have been shown to be predictors of training success (Saus, Johnsen, Eid & Thayer, 2012; Flin, 2001). The focus of this thesis is personality, as measured by the Five Factor Theory, cognitive constructs such as, Tolerance to Ambiguity and decision style and video game experience. These will be discussed in the following Sections.

2.3.2.1. Personality: One particular individual difference which has been considered when assessing confidence and accuracy in decision making is personality, as it can influence how people think, feel and behave (Roberts, 2009). One common theory and measurement of personality is the Five Factor Theory of Personality which consists of five dimensions of personality (Costa & McCrae, 1992). The five factors include: 1) Openness to experience which is characterised by a need for variety, novelty and change. High scorers tend to be more curious, imaginative, and excitable and individuals may show more interests in travel and have a variety of hobbies. 2) Conscientiousness refers to individuals who have a strong sense of achievement; high scorers tend to be more efficient, self-disciplined, organised and deliberative. 3) Agreeableness, is characterised by individuals who are more sympathetic, trusting and more likely to comply with others. 4) Extraversion is characterised by individuals who are sociable, adventurous and more enthusiastic and assertive in character. 5) Neuroticism is characterised by individuals who are easily upset, display higher levels of anxiety, irritably and have less self-confidence. Authors also characterise high neuroticism as having low emotional stability (Digman, 1990). With reference to the Five Factor Theory, there has been research which has shown that particular traits may relate to performance. It has been shown that high levels of neuroticism and low conscientiousness are linked to poorer performance (Mount, Barrick, Scullen & Rounds, 2005) as well as being valid predictors of job performance (Salgado, 1998). A meta-analysis carried out by Darr (2011) found that the personality traits of low neuroticism (emotional stability) and conscientiousness are likely to be core attributes for general military success. However, it has been further argued that the link between personality and performance is inconclusive (Saus et al., 2012). Research conducted on military personnel showed a weak correlation between personality and performance (Wallenius, Bäccman & Larsson, 2014).

Nevertheless, there is some research that has shown evidence to support the idea that personality traits may be related to decision accuracy and confidence. A study by Juanchich, Dewberry, Sirota and Narendran (2016) which investigated the predictive values of real-life decision outcomes, found that personality traits of extraversion and conscientiousness predicted decision outcome scores. Furthermore, conscientiousness was found to have a positive relationship with decision outcome whereas extraversion had a negative relationship on decision outcome. Additionally, Schaefer, Williams, Goodie and Campbell (2004) found support for personality playing a role in confidence judgements in decision making. The personality trait of narcissism, which is associated with an inflated self-view and proneness to take risky decisions, and extraversion significantly predicted overconfidence. Individuals with these traits were more likely to apply higher confidence levels. However, the confidence was not related to the accuracy of the decision. In comparison, the personality trait of openness to experience/intelligence was significantly related to both accuracy and confidence (but not overconfidence). Supporting this finding, Buratti, Allwood and Kleitman (2013) also found that openness predicted higher confidence levels. As mentioned in Chapter one, overconfidence can result when individuals apply higher levels of confidence to incorrect responses. To examine the relationship between confidence and accuracy, a recent study by Wheatcroft et al. (2017) assessed the suitability of unmanned aerial system (UAS) supervisors for operator selection. The authors found that neuroticism was negatively related to confidence, conscientiousness positively related to confidence and an intolerance of ambiguity was negatively related to W-S C-A. Consequently, research demonstrates that personality may be linked to the cognitive processes involved in decision accuracy, confidence and metacognitive ability in the air defence context.

In addition, personality has also been shown to be related to other concepts relevant to military decision making, such as stress, WL and Situational Awareness (SA) (Saus et al., 2012 – these will be discussed in greater detail in Chapter 3). Indeed, low scores on neuroticism and high scorers on extraversion and conscientiousness predicted subjective and observer related SA (Saus et al., 2012). These traits are often associated with a resilient personality type (Campbell-Sills, Cohan & Stein, 2006) which is defined by an individual's ability to cope with stressors (Connor & Davidson, 2003). Flin (2001) found that traits of low neuroticism and high conscientiousness have also been shown to be related to successful training of emergency service recruits. Furthermore, Saus et al. (2012) argue that personality may be a relevant measure for people working in high workload environments. The current thesis has therefore included measurements of personality to increase understanding of its relationship to performance as well as examining personality in relation to decision accuracy, confidence and metacognition in an air defence scenario.

2.3.2.2. Cognitive Constructs: As well as personality, individuals differ in terms of cognitive constructs. These have also shown to play a role in individual decision making. Research has demonstrated that participants with lower cognitive abilities i.e. intelligence, depend more on heuristics to assist them to make a decision than individuals with higher cognitive abilities (Gonzalez, 2004). Hence, individuals may use different decision making strategies when making decisions dependent on their cognitive ability. Furthermore, individuals also differ in their ability to deal with uncertainty. Constructs which relate to coping with uncertainty include Tolerance to Ambiguity and Need for Closure, which also combine to form a Decision Style. These constructs are strongly related to individual's ability to deal with uncertainty. Due to the nature of Ops room decision making, these constructs have been considered in the thesis. The precise measurements used in this thesis are discussed in more detail in Chapter four.

2.3.2.2.1. Tolerance to Ambiguity: Budner (1961) argues that individuals who are less tolerant find ambiguous situations threatening in comparison to those more tolerant who tend to view these situations as desirable. Consequently, the uncertainty of an Ops room may be problematic to individuals with a low Tolerance to Ambiguity. Research has also suggested that a Tolerance to Ambiguity may also play a role in decision accuracy and confidence. Although Ghosh and Ray (1997) found that a tolerance positively affected decision maker's confidence, the research holds mixed views on whether intolerance would lead to lower or greater decision confidence. McGhee, Shields and Birnberg (1978) found no differences in confidence in individuals that scored differently on Tolerance to Ambiguity scale. An explanation for higher levels of confidence could be related to the fact that individuals who have more Tolerance to Ambiguity believe they have more control of their environment and, as such, display higher levels of confidence. Tolerance to Ambiguity has also been shown to interact with other factors such as task complexity. Endres, Chowdhury and Milner (2009) found that in a highly complex task, individuals with higher levels of tolerance were found to be more accurate and displayed higher levels of self-efficiency (confidence). Additionally, Iannello, Mottini, Tirelli, Riva and Antonietti (2017) found Tolerance to Ambiguity to be a predictor of work related stress. Consequently, in high stress and WL environments, it would be beneficial to have an ability to tolerate ambiguity and uncertainty. Indeed Adams-White et al. (2018) found a positive relationship between Tolerance to Ambiguity and accuracy. Hence, a Tolerance to Ambiguity may be a beneficial personal characteristic to aid in decision making in air defence.

2.3.2.2.2. Need for Closure: The Need for Closure is "the desire to possess some knowledge on a given topic, any definite knowledge as opposed to confusion and ambiguity" (Mayseless & Kruglanski, 1987, p.164). The Need for Closure has been linked to individual differences in decision making. For instance, research has demonstrated that high scorers tend to reduce the amount of information they process (Webster & Kruglanski, 1994). In terms of decision accuracy, confidence and metacognition, the Need for Closure has been linked to high confidence as individuals make fewer hypotheses and display higher levels of confidence in their chosen hypothesis (Yang et al., 2012). However, it is important that this confidence is linked to the accuracy of the decision. As such, the Need for Closure will also be included in this thesis as a potential characteristic in aiding decision making in air defence.

2.3.2.3. Video Game Players (VGP): The role of an air defence operator is cognitively demanding. Operators must visually attend to the radar screen, communicate with other operators to make decisions and make difficult decisions in a high-paced, dynamic environment. Research has highlighted the potential benefits of investigating certain populations to increase understanding of individual differences in decision making (Wheatcroft et al., 2017). Research which has investigated the skill set of VGP has shown that the VGP population have been positively related to a range of sensory, perceptual and attentional skills (Spence & Feng, 2010) as well as high levels of visuo - spatial attention skills (McKinley, McIntire & Funke, 2011). Additionally, Lin et al. (2015) found VGPs to demonstrate lower distress and worry in relation to a simulated imaging and weapon release task. Hence, VGPs show superior performance under high cognitive demands. Moreover, a recent study showed that video gamers were better able to discriminate their accurate and inaccurate responses demonstrating higher metacognitive awareness. Wheatcroft et al. (2017) found that VGPs, in comparison to private and professional pilots, possessed skills which led them to be the least likely to exhibit overconfidence in decision judgements. Gaming has also been shown to be a useful predictor of performance. It has been argued that recent experience and/or task-specific experience are good predictors of performance in comparison to lifetime experience (Wheatcroft et al., 2017; Wiggins & O'hare 1995; Nicholson & O'Hare, 2014). This would suggest that those who play video games may perform better on a task due to the similarities of interacting with a computer system. Consequently, the use of different populations may also provide valuable insight into the difference in cognitions and decision making between Non-Gamers (NGs) and VGPs. Research could also increase the understanding of the relevance of task-specific experience related to decision making (Nicholson & O'Hare, 2014); hence, provide a unique insight into the transferability of skills and ability to the air defence domain.

VGP has also been related to WL and SA. Due to the higher cognitive abilities found in gamers, they are better able to cope with the demands of the task (Chandra et al., 2016). As such, there is potential to increase understandings of individual differences related to air defence decision making by examining the impact of VGPs on air defence decision making. Furthermore, in the light of NDM research, sub-populations such as VGPs may provide valuable insight into gaining expertise and the transition of skills. The research in this thesis aims to extend current research by Adams-White et al. (2018) by examining VGPs in an air defence decision making task as well as investigating the influence of the hours spent playing games, as this has shown to impact on operator performance (Lin et al., 2015).

2.4. Summary

Chapter two has introduced the theoretical backdrop to the thesis. As such an integrative approach has been applied in this thesis which combines elements of classical and NDM theories. Consequently, this Chapter has also provided the methodological backdrop for the development of the metacognitive measures used in this thesis to assess decision accuracy, confidence and metacognitive ability, as discussed in the following Chapter. Decision making is influenced by the environment, the task and the individual, This Chapter reviewed the external (TL, DC, time pressure, audio) and internal factors (personality, cognitive constructs, VGPs) that have the potential to relate to decision making and which will manipulated in the experiments conducted in this study. In summary, this Chapter has highlighted the need to understand air defence decision making as part of an interaction between the individual and the situation and, as such, the thesis has examined how these factors related to decision making in air defence

CHAPTER THREE

3. Methodological Approaches and Measurements

3.1. Introduction

The second aim of this thesis was to expand on previous measures of performance and confidence by examining the metacognitive ability of individuals in a critical environment via the application of a novel measure. As such, this project developed a low fidelity, yet realistic computer-generated experimental stimulus for the methodological approach which combined objective measures of accuracy, alongside subjective measures of confidence. The aim of the measures was to improve the understanding of air defence decision making and the metacognitive abilities of air defence personnel by combining elements of experimental laboratory testing with NDM methodology. In light of this, this Chapter is divided into two Sections. The first section reviews previous methods that have been used to investigate decision making in critical environments in the NDM domain. Where possible, this is discussed in relation to air defence. Second, the Chapter reviews some of the previous measures that have been used to study metacognition in other domain-specific environments. It introduces the method used in this thesis and talks to the validation of the method in relation to a published paper. The author contributed to this work, which used W-S C-A to analyse the suitability of UAS supervisors (see Wheatcroft et al., 2017).

3.2. NDM Methods Background

There has been a plethora of research that has sought to increase and understand decision making in air defence. Indeed, TADMUS (as discussed in Chapter 1) was set up in response to an air defence incident. Various methods have been used to investigate decision making in critical environments. As such, a continuum of methods exists, which range from laboratory-based to real-world settings. As discussed in Chapter two, these methods are generally seen as two separate research approaches, NDM and

CDM. However, more recently, there has been an emphasis on the potential to combine the strengths of both schools of thought to create an integrative theory (Markman, 2018). As also discussed in Chapter two the method in this thesis is more aligned to that of NDM but aims to integrate some of the methods from CDM theory. NDM research endeavours to be high in ecological validity and is very domain specific. However, more experimentally-based methods may be of benefit to NDM research (Lipshitz et al., 2001; Markman, 2018) as these methods can allow for more controlled testing to enhance the understanding of variables involved in the decision making process. In light of this, and the nature of the research domain of this project (i.e., air defence), this section will focus on reviewing methodologies to understand decision making in critical environments relative to the NDM paradigm together with the methodological integration of more CDM approaches.

Previous NDM research, which has examined cognition and decision making in critical environments, has used a range of different methods, all of which aim to gather a deeper understanding of decision making processes and how experts make decisions. These methods range from knowledge elicitation techniques i.e. questionnaires (Klein & Militello, 2001) and in-depth interviews (Kaempf, Klein, Thorden & Wolf, 1996) to high fidelity simulations (Calfee & Rowe, 2004). First, knowledge elicitation methods will be considered. Knowledge elicitation methods attempt to capture real-world decision making processes and include techniques such as interviews, questionnaires and Cognitive Task Analysis (CTA).

3.2.1. Cognitive Task Analysis (CTA): A widely used knowledge elicitation method in NDM is CTA, which is aimed at understanding the cognitive processes involved in expert decision making. Specifically, it aims to uncover the decision requirements that underpin an experienced person's job and/or task performance (Klein, 1996). Hence, it has been defined as "a set of methods to elicit, explain, and represent the mental processes involved in performing a task" (Klein & Militello, 2001, p. 168). Within CTA, there are also a wide range of methods which include interviews and observation methods, simulations and self-report. As such, different CTA methods aim to capture different aspects of expertise such as, mental modes, attention, perceptual skills, recognition of typicality, routines and strategies, and memory.

One method of conducting CTA is the Critical Decision Method (CDM) which is a semi-structured interview technique where individuals are required to recall a particular event (Crandall, Klein, Klein & Hoffman, 2006). During the interview, once the initial recall has occurred, sweeps or phases which further probe the individual follow. These sweeps aim to identify decision points, generate timelines, gain a deeper understanding of the situation and the final sweep probes include asking the individual for "what – if" queries. Hence, CTA allows for a deeper understanding of an event as verbalised by an expert. The outcome is a detailed and specific account of an event which provides extensive knowledge on the decision making process (for a full review see Hoffman, Crandall & Shadbolt, 1998).

With regards to air defence decision making research, Kaempf et al. (1996) used CDM to analyse the decision strategies of air defence operators. Through the use of CDM, the authors found that decision makers tend to use recognition processes when making decisions. That is to say that their decisions were based on previous experiences in similar situations. Further, the decision makers also tended to use feature matching and story building to help with their decision making. This study also highlighted the importance of situational awareness in decision making. Hence, the use of CDM can be beneficial in generating an understanding of a domain-specific event and how experts use their experience to assist them in their decision making. Moreover, the outcomes from CTA may also be used to develop and investigate decision making processes further.

CTA can also be used with other research methods such as decision ladders and the ShadowBox technique. Decision ladders develop prototypical models of activity (Rasmussen, Pejtersen, Goodstein, 1994) and have been used to assist with the design of new technology and systems (Salmon, Jenkins, Stanton & Walker, 2010). More recently CTA has been used to develop the ShadowBox technique (Klein, Hintze & Saab, 2013).

The ShadowBox technique aims to assist novices in understanding the decision making processes of experts. This is then used to compare expert decisions alongside novices. Hence, CTA provides a good basis for collecting qualitatively rich data.

However, the process of CTA can be time-consuming with each interview taking up to 3 hours to conduct with trained interviewers. To counter this limitation, the Applied Cognitive Task Analysis (ACTA - Militello & Hutton, 1998) was developed. ACTA aims to help gain critical cognitive elements from SMEs. ACTA is divided into 3 techniques, 1) task diagram interview, 2) knowledge audit and 3) a simulation interview. First, the task diagram interview aims to provide a broad overview of the task and to identify any cognitive complex elements of the task. The purpose of the knowledge audit is to provide detail and examples of cognitive elements of expertise contrasting with expert and novices. The simulation interview generates specific, detailed information about an expert's cognitive processes within the context of a challenging scenario. The expert is presented with a scenario and during the simulation, they would be asked to identify major events, including judgements and decisions. (Millitello & Hutton,1998). Although these techniques used in CTA provide valuable insights into expert individual decision making, there are some limitations to this method. For instance, there are difficulties with retrospective verbal recall and biases as individuals are thinking back on a past events (Nisbett & Wilson, 1977). In addition, due to the specificity of the domains, the sample size tend to be relatively small and, due to the amount of information generated from them it is only one event is generally covered. On the other hand, scenario-based methods are, relatively speaking, quicker to conduct and information can be gathered on across a range of different scenarios. These methods are reviewed in what follows.

3.2.2. Situational Judgements Tests (SJTs): SJTs are scenario-based, knowledge elicitation methods and require participants to evaluate a course of action for the likelihood that they would perform the action and for the effectiveness of an action (Sorrel et al., 2016). In relation to understanding decision making in critical environments, it has been argued that SJTs show predictive validity for constructs such as, knowledge and skills, applied social skills, basic personality tendencies and heterogeneous composites (Christian, Edwards & Bradley, 2010). SJTs have also demonstrated to be effective predictors of job performance (Christian et al., 2010) and individual differences (Reinerman-Jones, Matthews, Burke, Scribner, 2016). However, SJTs may lack external validity.

3.2.3. Tactical Decision Making Games (TDGs): Another way in which decision making has been investigated in critical environments is through the use of War Game experiments and TDGs. These are low-fidelity training techniques used to understand and improve tactical skill and decision making ability in the military (Gonsalves, 1997). TDGs were originally developed by US marines. TDGs involve a short written scenario with decision points and a sketch to show graphics or a map. These scenarios usually take in the region of 10 minutes to complete. The scenarios are facilitated by trainers and, once completed, the outcomes are discussed. One of the aims of TDGs is to help teach individuals how to think. Experts can implicitly communicate their thought processes to less experienced personnel. The benefits of TDGs include the generation of qualitative data that can be analysed to understand individual decision making processes. However, similarly to the previous methods, these rely on expertise with already accumulated tacit knowledge. Once again, therefore, there is reliance on SMEs being present. Nevertheless, the act of implicitly communicating thought processes has also been used through the use of think-aloud protocols.

3.2.4. Think-Aloud Protocols: Think-aloud protocols have been demonstrated to be a valuable tool for investigating decision making strategies and assisting with training of critical thinking in military decision making (Cohen, Freeman, & Thompson, 1998). To counter some of the limitations of CTA and interviews, think-aloud protocols gather qualitative data on thinking processes and behaviours during a concurrent task. In general, they require the individual to talk through their decision processes as they go through the task. Frye and Wearing (2014) investigated decision making in bush fire-fighters using this method. The authors found that expert bush fire-fighters use previous experience to assist them with, what the authors refer to as, a metacognition loop (i.e., monitor, decide, and act). Such a finding is similar to that of Cohen et al. (1998) who argued that decision makers use pattern recognition to support metacognitive skills. The benefits of think-aloud protocols include a reduction in the chance of memory delay as compared to more retrospective methods of questionnaires and interviews which ask individuals to think back on a specific event or scenario. As such, these methods have been used to identify metacognitive processes and develop models to understand decision making.

Knowledge elicitation techniques focus on producing extensive qualitative rich data in order to understand the decision making process. The use of experts in these techniques provides high ecological validity. However, there are difficulties in gaining access to suitable SMEs which limits the potential sample size available. In addition, conducting high-fidelity research in training centres and real ship exercises is expensive and time-consuming with limited access available to researchers. A further critique in is the lack of scientific rigour due to being unable to control for variables, and thus reducing internal validity (Markman, 2018).

3.2.5. Simulations and *Microworlds*: One methodology which has been used to bridge the gap between experimental control and fidelity to investigate decision making in critical environments is the use of *microworlds* (Brehmer & Dorner, 1993; Gray, 2002). Fidelity is an umbrella term. It has been argued that fidelity usually falls between two types, physical and psychological fidelity (Liu et al., 2009). Physical fidelity has been defined as the degree of similarity between the real and simulated environments (Allen, Hays & Buffardi, 1986). In comparison, psychological fidelity is concerned with similarities of the psychological and cognitive constructs. Psychological fidelity involves using psychological theory and processes to direct the design of the simulation. Therefore, it can be argued that increased psychological fidelity over physical fidelity aims to understand more fully psychological constructs and cognitive mechanisms. Hence, high-fidelity is not necessary for skill transference (Dahlstorm, Dekker, Van Winsen & Nyce, 2009). *Microworld*s and simulations can be beneficial in helping to improve the understanding of air defence decision making as well as to meet training needs.

One way physical fidelity has been introduced to decision making is through the use of *Microworlds*. *Microworlds* are simulated task environments first introduced by Turkle (1984) and are generally computer-based behavioural simulations which aim to assess dynamic decision making (DDM - Brehmer & Dorner, 1993). DDM has been defined as “interdependent decisions made in an environment that changes as a function of the decision sequence, or in both ways” (Gonzalez et al., 2005, p. 273) Hence, high physical fidelity *Microworlds* increase ecological validity and examine the decision making processes in a more experimental manner.

Furthermore, simulations offer an alternative to field work and interviews by addressing some of the criteria for NDM as described by Orasanu et al. (1993). Simulations allow researchers to test dynamic, continually adapting and changing environments in a controlled way. The level of fidelity varies, with some research aiming to achieve high levels of physical fidelity. For instance, the AEGIS Cruiser Air-Defence Commander

(ADC) is a high fidelity simulation which models all aspects of the ship and was built to model the performance of U.S. Navy personnel known as Watchstanders engaged in air defence (Calfee & Rowe, 2004). The aim of the ADC was to uncover the cognitive aspects of naval air defence and model decision making. The body of work conducted using the ADC provided insights into a wide range of aspects of decision making in air defence through the use of high fidelity simulations. This included Watchstander skill experience, fatigue, type of decision making and environmental influence of the performance of the individual, as well as the team. The outcomes of the performance are logged in order to assess the findings and help with future training by allowing the investigation of a range of factors related to decision making. Similarly, Liebhaber and Smith (2000) used a *microworld* to identify and describe factors and cognitive processes that an air defence team uses to assess and prioritise aircraft contacts. Six Navy officers took part in this study. The participants had to watch and analyse six tracks which were characterised by a range of different threat levels and different track types. In total 22, factors were identified. The most important being signal emissions, course, speed, altitude, point of origin, IFF responses, flight profile, intelligence information, and distance from the detector. Hence, research conducted in simulators and *microworlds* provide extensive data based on high-fidelity research.

There has been some debate about the use of simulations and *microworlds* to understand decisions in highly critical and dynamic environments. For instance, Chapman Nettelbeck, Welsh and Mills (2006) argue that there are problems with simulations' construct validity. In a fire-fighting *microworld,* Chapman et al. (2006) found no difference between experienced and non - experienced participants. Thus, the authors argued that simulations do not produce results that are generalisable to real-world decision making. Nevertheless, Elliott, Welsh, and Mills (2007) argue that simulations can be useful in adding to the understanding of psychological processes. Elliott et al. (2007) investigated whether concepts in a *microworld* related to NDM. It was found that, although their use may be limited, there were findings to suggest that the perceptual-cognitive skills were similar to that of those observed in experts. This suggests that simulations are useful in understanding psychological constructs via increased psychological fidelity (Kozlowski & Deshon, 2004).

Although this thesis did not develop a high physical fidelity *microworld,* the measure and method was informed by the benefits of *microworld* simulations and was adapted to create the method used. The aim of the method was to isolate variables of interest but still maintain the complexity and dynamics of real-world decision making (Markman, 2018). Part of the TADMUS research discussed in Chapter one, was the development of research methods (Johnston, Poirier & Smith-Jentsch, 1998). Johnston et al.'s (1998) suggestions included; the importance of psychological fidelity and the assistance of SMEs to develop quasi-experimental. Furthermore, Johnston et al. (1988) concluded, that the criteria for research methodology include a) creates acceptable level of fidelity, b) enables opportunities for assessing individual and team performance processes and outcomes, c) supports research designs for testing the impact of training interventions on performance under stress, and d) include Navy trainees and ships teams as research participants. As such, the method applied in this thesis aimed to follow this guidance.

However, gaining access to experts in this domain, as well as obtaining sufficient participant numbers to conduct experimental work is difficult due to their availability. In addition, the PWO role is highly specialised. As such, in ideal circumstances, this method would have been carried out fully with expert decision makers. However, due to numbers and access available to the author, it was decided that novices would also take part. Nevertheless, as suggested by (Hoffman & Klein, 2017), it may be beneficial to the NDM paradigm to gain an understanding of how expertise is developed. For instance, Klein et al. (2013) used the ShadowBox method to help novices understand the decision making processes of experts. Therefore, the use of novices in this study can also be considered positively as it allows for a baseline comparison to help understand decision making in critical environments. Further, the use of novices and different populations, such as

VGPs may be of benefit in understanding the training needs of less experienced decision makers.

3.3. Measures of Metacognition

This thesis examines metacognitive abilities of air defence operators which was achieved by assessing the relationship between decision confidence and accuracy using the W-S C-A measure. There have been various ways in which metacognitive ability has been measured previously and the next section outlines and reviews some of that previous literature.

3.3.1. Signal Detection: One of the earliest measures of metacognition which quantifies decisions using analytical methods is the Signal Detection Theory (SDT-Tanner & Swets, 1954). Type 1 SDT refers to the classification of the stimulus and Type 2 SDT assesses individual confidence in correct and incorrect responses (Clarke, Birdsall & Tanner, 1959). As previously discussed in Chapter one, decision confidence is a useful way of measuring metacognitive ability. SDT distinguishes between *sensitivity*, this relates to individual's ability to discriminate stimuli and *response bias*, this is an individual's response strategy for dealing with ambiguous stimuli. It is used in ambiguous situations such as where a "noise" present or not present. The outcomes include a Hit (correct response, correct confidence), Miss (incorrect response, incorrect confidence), False Alarm (correct response, incorrect confidence) or a Correct Rejection (incorrect response, correct confidence).

SDT is traditionally applied to psychological perception studies to understand perception, memory, and human vigilance. However, there have been studies which have applied SDT to other domains such as eyewitness testimony (Brewer & Wells, 2006). Furthermore, in a military setting, Eubanks & Killen (1983) used SDT to understand changes in pilot decision making behaviour. It was concluded that SDT is beneficial for training and can provide a conceptual framework for evaluating training.

3.3.2. Self-reported: Self-reported measures, such as questionnaires have been used to measure metacognitive ability. These questionnaires include the metacognitive awareness inventory (MAI - Schraw & Dennison, 1994), which was designed to assess knowledge of cognition and regulation of cognition. The MAI is a self-report scale consisting of 52- inventory items. Limitations, include potential reporting biases in self-report measures and as such, these may not an accurate measurement of metacognition. Furthermore, it may be beneficial to introduce more quantitative and numerical measures to understand metacognition in critical environments by measuring subjective feelings alongside objective measures.

3.3.3. Calibration: Another way of measuring metacognition is through the use of calibration. Calibration assesses actual accuracy and perceived accuracy. Hence, allowing individual's awareness of the accuracy or inaccuracy of their decisions to be measured. A poor calibration would demonstrate no relationship between actual and perceived performance or displays over or under confidence in performance. This is commonly assessed through decision confidence (discussed in Chapter 1; Jackson & Kleitman, 2014; Schraw, 2009). Decision confidence has been measured in various ways which include, as a percentage from 0% = *guessing* to 100% = *absolutely certain* they are correct, as well as the use of Likert scales. Measurements in this manner then generate biases which assess how well the individual has matched their accuracy and confidence (Stankov, Morony, Lee, Luo & Hogan, 2012). Jackson and Kleitman (2014) developed a Medical Decision Making Test (MDMT). Consequently, this method has been used in a wide range of domains to assess performance accuracy as well as SA.

The calibration method has been previously used to examine meta-SA. Meta-SA refers to an individual's confidence in self-ability to discriminate between true and false descriptions of the situation (Lee, 1999). Meta-SA is a useful measure, as it is thought not only to be representative of the characteristic of the operator (Keren, 1991), but can also be helpful in explaining human performance (Lee, 1999). Results have found that meta-SA changes over time (Lichacz, 2008) and individuals also display overconfidence in SA assessment (Sulistyawati, Wickens & Chui, 2011). Additionally, SA may also impact on decision making. For instance, individuals displaying overconfidence in their SA ability may be more likely to stop searching for more information to aid prediction. This may lead to higher levels of risk taking resulting in poorer decisions. However, this aspect was not specifically investigated by Sulistyawati et al. (2011). A study by McGuinness (2004) used a calibration method to assess Meta-SA and developed a Quantitative Analysis of Situational Awareness (QUASA) measure. In which individuals answered probe statements regarding SA and were then asked to state how confident they were in their assessment. These studies indicate that an appropriate SA-confidence calibration could lead to an appropriate decision being made. However, SA is limited in providing information about actual decision making. It cannot be said that SA-confidence calibration would successfully translate into a good decision being made, as many factors may influence the prospect of good SA and its relationship with successful performance (Stanners & French, 2005). Hence, understanding how W-S C-A factors in other performance measures such as WL and SA may generate deeper understanding about effective decision making. These studies also provide support for understanding decision making through a calibration style methodology which allows for a broader understanding of the situation compared to other measures or raw scores (Lichacz, 2008).

3.3.4. Within-Subjects Confidence-Accuracy: W-S C-A is a measure of metacognition and has been defined as a "calculation which enables expression of individual confidence in each incorrect or correct response made" (Wheatcroft & Woods, 2010; p.195). Simply put, the W-S C-A method is able to calculate the statistical relationship between the levels of confidence individuals might place in responses given relative to the corresponding decision correctness. Point bi-serial correlations are used to assess the individual C-A relationship. Higher scores indicate an appropriate level of confidence to a response. For example, high confidence is applied to a correct response and lower confidence to an incorrect response. Negative scores indicate higher confidence in incorrect responses or low confidence in correct scores. W-S C-A has been used successfully in domains such as forensic, investigative and legal psychology (Wheatcroft &

Woods, 2010; Wheatcroft et al., 2004). More recently W-S C-A has been used to examine the suitability of supervisory personnel for UAS; Wheatcroft, et al. (2017). The validation of the method is discussed in Section 3.5. Hence, there is potential that it may be applicable as a performance measure to more critical environments such as air defence.

3.4. Human-Machine Interaction and Performance Measurements

Performance measures have been used to assess decision making in critical environments. As previously discussed, and the most common and important within the military environment are SA and WL (St John, Callan, Proctor & Holste, 2000). These also often tend to be considered in relation to each other (Endsley, 1995, Vidulich & Tsang, 2012; Wickens, 2002). However, there is some scepticism as to how much these concepts stand up to scientific rigour (Dekker, Hummerdal & Smith, 2010). Nonetheless, Parasuraman, Sheridan and Wickens (2008) argue that these constructs are both predictive of performance and diagnostic of operators' state, allowing judgements to be made surrounding their overall performance. Furthermore, the inclusion of these measures enables the combination of multiple performance measures and can offer a complete representation of operator experience (Ikuma, Harvey, Taylor & Handal, 2014). Additionally, it enables the examination of how the measure of W-S C-A aligns with the wider measurements currently used in human-machine interaction decision making literature.

3.4.1. WL: Workload is multi-dimensional and has been described as "the relation between the function related to the mental resources demanded by a task and those resources available to be supplied by the human operator" (Parasuraman et al., 2008; p.145). WL can be increased by time constraints, the amount of information available, whether concurrent tasks must be conducted etc. which impact on performance. Higher levels of WL may, therefore, impede individuals to make accurate and confident decisions. WL was found to be negatively related to overall decision confidence. Adams-White et al. (2018) found that higher levels of reported WL reduced decision confidence scores. This has important implications for air defence decision making, as reduced confidence in any decisions could also lead to increased WL, as individuals seek out more information to support or contradict the certainty of their decisions. With regards to metacognition, Kim, Macht and Li (2012) investigated whether there was a relationship between WL and metacognition. Results from this anti-air warfare simulation task found that individuals had a negative correlation between metacognition and WL. Thus, increases in WL in the task may impair metacognitive ability. Research, therefore, implies that high WL negatively impacts on accuracy, confidence and metacognition. However, little is known on other factors which may mediate this relationship which includes the external and internal factors discussed in Chapter two.

WL is generally a construct related to individual's resource capacity and the impact of the stressor on the individual (Friedrich et al., 2018). Individual differences have been demonstrated in the subjective experience of WL. As such, WL has been seen as an interaction between the characteristics of the person as well as the task (Szalma, 2009; Chiorri, Garbarino, Bracco & Magnavita, 2015). In particular, Adams-White et al. (2018) found WL to be negatively related to openness to experience. This construct refers to an individual's preference for novelty and curiosity. As such, high scorers are reported to be more imaginative and broad-minded (Costa & McCrae, 1992). It also supports previous findings that higher openness is related to a greater resilience to stress in comparison to low scorers on this scale (Williams, Rau, Cribbet & Gunn, 2009) as well as being negatively related to perceived situational demands (Penly & Tomoka, 2002). There is also research to suggest that VGPs have the ability to deal with higher levels of WL. By including VGPs in this research, it is envisioned that particular skills could be found which may relate to WL. Indeed, research has shown that video gaming can increase attentional resources (Boot et al., 2008) and attentional visual field (Hubert-Wallander, Green & Bavelier, 2011). Gonzalez, (2004) argued that high WL is more detrimental in individuals with low cognitive abilities than high. Hence, VGPs may be better suited to the demands of the task than NGs.

3.4.2. SA: The most frequent and widely accepted definition of SA is provided by Endsley (1988). Through her work in aviation, Endsley describes SA as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1988; p.792). Within this description, three elements of SA are prescribed; comprehension, perception and projection. Level 1 SA is concerned with the perception of elements in the environment, level 2 SA is the comprehension of their meaning and level 3 SA requires the projection of their future status.

SA is believed by many as to be integral to any military decision making (St John et al., 2000). Furthermore, SA is required to make complex decisions. A higher level of SA leads to faster and better decision making (Endsley & Jones, 1997; Rouse, Cannon-Bowers & Salas, 1992). However, individuals can have good knowledge of the situation but not necessarily make the correct decisions. Likewise, a good decision can be made with poor SA (Endsley, 1995). Hence, research into the direct relationship of SA on the decision making process is warranted, particularly as much of the previous research suggests only a probabilistic relationship; thereby, assumed - not guaranteed. A study which investigated the assumption that good SA is related to good decision choices was conducted by Stanners and French (2005). Although the authors found a positive correlation between SA and decision making; that is to say, high levels of SA were related to high-quality decision making, the findings were only of a medium strength. This would suggest that other factors may be involved in turning good SA into a successful performance i.e. correct decision. Hence, there might be more to good decision making than having a good assessment of the situation, other measures may be useful to understand decisions made within the context of human-machine interaction in military environments.

As discussed in Chapter two, there are also some individual differences in SA which are important to decision making (Saus et al., 2012). For instance, SA has also been linked to confidence. Adams-White et al. (2018) found that SA was related to decision confidence. Individuals who reported higher levels of SA were also more confident in their decisions. However, as SA was only recorded subjectively, this finding could also suggest that decision confidence relates to confidence in SA judgements. Thus, these findings should be regarded with caution, as objective SA was not considered and a confidence bias has been found in SA reporting (Sulistyawati & Chui, 2009). Thus, it might be that individuals are generally confident in their assessments of SA performance. Importantly though, SA was not related to accuracy in decisions taken and thereby individuals may privately believe they had a better understanding of the situation than they accepted. Supported by Endlsey (1995), self-ratings tend to be related to a statement of how certain the person feels about SA. Adams-White et al. (2018) found no relationship between SA and W-S C-A. However, there is a dearth of research which has specifically investigated video game experience and the use as gamers as a population in this domain before.

In light of the literature presented, both WL and SA will be considered in the research to assess their role in individual's decision accuracy, confidence and metacognition. WL in the current research was assessed as a function of TL, as increases in the amount of information (TL) increases WL. To examine whether this was the case, WL was provided as a manipulation check to see if WL varied under different TL conditions. Further, as assessments of the manipulations, both WL and SA have been found previously to be sensitive to variations in TL. Hence, changes in TL are a function of both WL and SA (Selcon et al., 1991).

3.5. Validation of Method

The next section of this Chapter presents a study in which the author used the W-S C-A methodology to validate the approach in this thesis. Wheatcroft et al. (2017) used a similar method to assess the suitability of a UAS supervisory role. Specifically, different populations were compared to assess the most metacognitively confidence - accuracy sensitive group in a simulated civilian cargo flight task. In the published paper, the author contributed to the work by conducting further analysis on the data to investigate the impact on automation versus manual decisions in individual's decision confidence and accuracy and metacognitive sensitivity. Wheatcroft et al. (2017) examined the suitability of UAS supervisors by examining the impact of decision danger, where danger was associated with decisions carrying more risk, and different populations which included professional pilots, private pilots, VGPs and a control group. Questionnaires were also provided to examine personality constructs. In the study, a series of pre-recorded video vignettes of typical scenarios that might be encountered during a typical flight was displayed to participants. A decision log was developed, which included 21 decision events and each event had 3 decision options for the participants to choose from. Included in these options, there was always the option to allow the autonomous system to control the UAS and intervene and manually fly the vehicle. Of the options, one was considered to be the correct response which allowed for an objective measure of decision accuracy. Once an option had been selected, participants were required to rate how confident they were in their decisions, providing a subjective measure of decision confidence. These measurements then generated an individual W-S C-A calculation for each participant.

As discussed in Chapter two, the findings from this study which utilised the measure showed that decision accuracy and confidence was influenced by decision danger. In addition, this study demonstrated that individual differences may play a role in decision accuracy, confidence and metacognitive ability. Furthermore, differences were also found between the populations. Decision confidence was found to be higher in professional pilots and also higher in VGPs than the control group. Hence, these findings demonstrate that there are both external and internal factors which are influential on individual decision accuracy, confidence and metacognition., the outcomes imply that W-S C-A could be a suitable performance measure to increase understanding of metacognition and decision making in more critical environments, such as air defence. In addition and as previously discussed in Chapter two, the thesis has included variables DC and TL, WL and SA as these variables in particular have been demonstrated to be relevant in air defence decision making (see Adams-White et al., 2018).

3.6. Metacognitive Feedback and Training

To recall, metacognition includes both metacognitive awareness and regulation. Metacognitive awareness involves the monitoring and control of one’s own performance and regulation which assessment of owns ability (Flavell, 1979). Poor metacognitive ability in an individual may, therefore, suggest that the individuals may lack an awareness of their understanding of a task, as well as limit their abilities to maximise performance. Indeed, the ability to monitor one’s own performance has successfully been linked to improved learning in an educational setting (Tanner, 2012). For example, if a student is aware that they have a limited understanding of a topic, they may seek help to improve their understanding. In contrast, individuals with poor metacognitive awareness may not know when to seek assistance, which, in turn, limits their learning and performance.

Furthermore, as noted in Chapter two, individuals often use cognitive heuristics to assist them in their decision making. However, there are also often problems with the use of these heuristics which lead to poor decision making, which is heightened in complex and uncertain environments (DiBonaventura & Chapman, 2008). One way in which researchers have aimed to reduce these problems derives from de-biasing. A method which has been found to been effective in reducing biases is the use of video games and training (Morewedge et al., 2015). Research has demonstrated that a training intervention, which included a video defining and explaining what heuristics were, had a de-biasing effect in both a range of context and across different forms of biases. For instance, Morewedge et al. (2015) found that games that produce personalised feedback and practice had both short term and long term effects on reducing bias. However, these methods tend to focus more of cognitive biases, and not necessarily the reduction of inappropriate confidence, related to metacognition.

As mentioned in Chapter three, Cohen et al. (1998) investigated metacognitive ability improvement using critical thinking. As such, training in critical skills included training in confirmation bias or fundamental attribution bias. Strategies were explained in how to recognise situations that may cause bias and how to prevent bias. One aspect of the training included evaluating their confidence in assessments. Individuals were asked to assume that their assumption is incorrect. As such, critical thinking skills, which enable individuals to consider confidence, may be an important consideration for training in decision making. However, training in this manner was found to increase confidence in assessments, but this was not necessarily related to the accuracy of these assessments.

More specifically, another way of improving metacognitive skill is via metacognitive prompting and metacognitive feedback (Hoffman & Spatariu, 2008; Fiore & Vogel-Walcutt, 2010). Hoffman and Spatariu (2008) define metacognitive prompting as "an externally generated stimulus that activates reflective cognition or evokes strategy use with the objective of enhancing learning" (p.878). This could take the form of questioning. Research has shown that metacognitive prompting increases decision making performance (Fiorella, Vogel-Walcutt & Fiore, 2012). On the other hand, metacognitive feedback provides information regarding individual's understanding or performance (Fiorella et al., 2012). Recent research has also begun to examine metacognitive monitoring in more critical domains such as air defence using high fidelity simulations (Fiorella et al., 2012). For instance, Kim (2018) examined the use of confidence de-biasing through feedback during a computer-simulated military-training task. Feedback was provided by giving participants information which compared their self-judgement to the actual response. This was visually displayed to the participants via an indication of under or over confidence. This study demonstrated that feedback has a positive training impact and improved individual's metacognitive judgements in some of the tasks. As such, there are benefits in training individuals to be more aware of the relationship between confidence and accuracy.

Nevertheless, there is insufficient research that seeks to understand the potential external and internal factors which may relate to individuals improvement in metacognitive monitoring and the use of other methods in this domain. For instance, research into personality has demonstrated a link between some personality types and better training success (Flin, 2001). Subsequently, there may be individual differences in respect of the benefits metacognitive feedback can provide (Dunlosky & Metcalfe, 2008). Jøsok et al. (2016) argued that metacognitive ability is a relatively stable construct of personality that can be quantified and made subject to training and improvement. It has also been argued that the cognitive and situational factors can determine the effects of training in a range of difference biases (Poos, Van den Vosch & Janssen, 2017). In addition, the improvement of metacognitive ability may also influence individuals SA and WL. Indeed Cohen et al. (1998) found that training in critical skills improved SA. However little is known about the influence of feedback on WL. A recent study by Kim (2018) found that metacognition did not increase WL in participants. Subsequently, it is therefore important that W-S C-A might also be applied as a way of assessing metacognitive ability and could, therefore, also be used in training.

3.8. Summary

There are a wide range of different methods and measurements which aim to investigate and understand decision making in critical environments such as air defence. This thesis has adopted an integrative approach between methods to develop a metacognitive performance measure. As demonstrated, it has been previously argued that NDM research should use and could benefit from a mixture of measures to reduce the limitations of using a single methodology (Lipshitz et al., 2001; Markman, 2018). Henceforth, and with the purpose of using elements of NDM research methodologies, the aim was to design and develop a method which would combine a more quantitative way of measuring metacognition in critical environments with naturalistic methodologies through the use of an SME generated scenario-based decision making task. As such, it is envisaged that the proposed method will provide a wider view of metacognition by examining the metacognition in a quantitative way in critical decision making environments. This will allow predictions about decision making to be tested experimentally. Chapter four will discuss, in detail, the Phases involved in the scenario development.

3.9. Conclusions

Both external and internal factors may potentially impact on decision making. In light of this, these factors have been examined in an air defence domain in relation to decision accuracy and confidence and metacognitive ability. The current thesis considers the impact of external factors such as TL, DC, time pressure and audio communications. What's more, what is lacking in the literature is a more individual focus on decision making and how metacognitive ability can be measured in these environments. Hence, internal factors which include individual differences (such as personality), cognitive constructs and video game play are examined. In addition, Chapter three discussed ways of measuring metacognitive ability and the integration of NDM theories and more traditional approaches to develop a method to examine metacognition in the air defence context. By combining these approaches, and using both subjective and objective measures to inform and measure metacognition, the work adds to understanding air defence decision making

In light of the above, and to address the aims of the thesis as set out in Chapter one, the work contained in this thesis addresses the following four research questions:

Research Questions

1) What are the external factors that influence decision confidence and accuracy in an air defence decision making task? Specifically, TL, DC, time pressure and audio communications, and what relationships exist with metacognition?

2) What, if any, are the individual differences involved in air defence decision making and how do they relate to metacognitive skills. For example, in light of personality, cognitive constructs and video game play?

3) How does the method and measurements used in this thesis align with the wider methods and measurements currently used to assess performance in decision making?

4) Can metacognitive ability in air defence be improved through feedback training?

CHAPTER FOUR

4. Experimental Method and Data Collection

4.1. Introduction

This Section of the Chapter will describe the steps involved in the development of the stimulus used in the Experimental Work. As discussed in Chapter three, the method used in the research studies aimed to combine some of the naturalistic elements of air defence decision making, such as uncertainty and time-pressure, with some of the constraints and rigour of more conventional laboratory testing, such as the use of independent and dependent variables. The method made use of realistic displays and scenarios with decision points which have been developed with, and verified by, Subject Matter Experts (SMEs). A significant amount of effort was taken to produce the experimental scenarios and corresponding materials involved in the studies reported. To gain the required skills to generate the computer simulations, the author attended lectures and workshops to acquire the underpinning knowledge required. These skills include using mathematical coding environments such as MATLAB® / Simulink as well as Virtual Avionics Prototyping Software (VAPS XT). It was also important to gain a strong understanding of the tasks and performance specifications required in air defence decision making with the help of SMEs.

The generation of the experimental stimuli and scenario took over 12 months to formulate, plan, initiate, test and finalise. Each Phase is discussed in turn: (1) SMEs; (2) scenario development; (3) computer-generated scenarios; (4) pilot study and (5) completed SME validated scenarios and decision logs. This Chapter also includes a description of the experimental setup and materials.

4.2. Stimulus Development

4.2.1. Phase One: Subject Matter Experts (SMEs):

The first phase of the scenario development involved meeting SMEs. The use of SMEs to assist in the experimental design is highly beneficial as they are able to provide a unique insight into the appropriate and relevant situations that are likely to be met and applied in the study context. Further, van den Heuvel, Alison & Crego (2012) argue that the utilization of SMEs to produce a "gold standard" of decision making can be useful in increasing the objective evaluation of decision effectiveness. As a result, SMEs were extensively used to help increase understanding of the domain for the author and provide some experimental content such as the decision logs and scenario design.

In this project three (3) SMEs with extensive knowledge of naval warfare, including roles as AWO and PWO positions were employed to acquire the domain-specific knowledge needed to provide the optimum, ecologically valid options for the task. Each SME has over 30 years of experience in the RN. At the time that the scenario was being developed, the SMEs worked for Systems Engineering and Assessments (SEA) Ltd. Their roles ranged from maritime team leaders to senior and principal consultants. In-depth meetings and discussions took place over two days to elicit the SME's knowledge with respect to air defence scenarios. A ship visit to HMS Daring was also organised. This allowed the author to witness an Ops room training session that demonstrated the execution of an air defence mission. Witnessing such an event in context allowed the author to develop a far more effective understanding of the Ops room environment, layout and how communications and decisions are made between and by individuals. This assisted in the scenario development. The ship visit also provided insight into the displays used and the layout of the radar screen in the ship space. The SMEs worked with the author to generate scenario-specific information and to identify response options to specific scenario events. This included the generation of the mission briefing and a realistic scenario in which the TL could be varied. Alongside this, an appropriate range of classifiable decision options, which varied in DC, were also established. To start the development of the scenarios, SMEs generated the image shown in Figure 1, which shows the original scenario design. It was decided by SMEs that a single peace enforcement scenario (see Appendix 8b.1) would be most the appropriate to use, where the TL could be varied to include three separate TL scenarios (high, moderate, low). This allowed the TL conditions to vary without over-complicating the scenarios and increasing the number of controlled and isolated variables. The criticality of each decision was also identified by the SMEs. A decision event with a higher consequence if made incorrectly held a high DC level. These divided across high, medium, low decision criticality events. DC is described in more detail in section 4.2.2.2. Descriptions of the differences in DC are also provided in Table 3.

4.2.1.1. Peace Enforcement Description: The peace enforcement scenario was developed with the assistance of the SMEs. It depicts a fictitious air defence task in which the participants are asked to take on the role of a PWO who has been tasked with monitoring and enforcing a NoFly Zone under the instruction of the United Nations Security Council Resolution (UNSCR- a full description is provided in Appendix 8b.1).

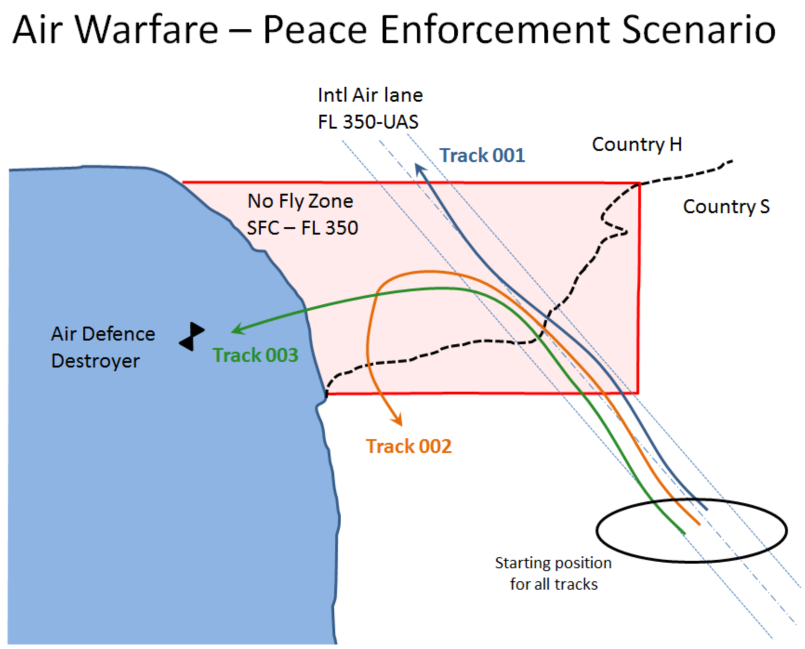


Figure 1. Schematic of the radar screen.

This figure displays the starting point for the aircraft track and the different routes they make over the course of the scenario. In Figure 1, Track 001 adheres to the airlane throughout the sceanrio, Track 002 has the same initial profile as as Track 001. However, it begins to head towards the ship's location and then back out of the NoFlyZone again. Track 003, again, starts with the same profile as Track 001 and 002. However, it continues to heads towards the ship. The decision events that occur in the scenarios are based on these tracks. The frequency and order of the Tracks varies depending on the TL condition. The figure also displays the additional features displayed on the screen such as the location of the NoFlyZone, the border between Country H and S and the ship's location.

4.2.2. Phase Two: Scenario Development

4.2.2.1. Task Load (TL): TL in the experiments corresponds to the number of tracks displayed on the radar screen and the frequency and speed at which decision events occur. In the low TL, decisions were based on one track being present on the radar screen at a time, a reduced frequency of decision events and longer periods of inaction (see Figure 2). In comparison, the high TL condition involved multiple aircraft tracks for the PWO to monitor on the screen at any one time, increased frequency of decision events and the requirement to make more than one decision at a decision event (see Figure 4). Figures 2, 3 and 4 display what is taking place seven minutes into the scenarios on each stress condition to highlight the differences in TL.

These images have been recreated in schematic form for clarity. For reference to the symbology displayed in these figures please refer to symbology key in Section 4.2.3.5. For the screen shots of the actual task scenario please see Appendix 7. Depending on the TL conditions, the sequence in which the decision events occurred varied. This was due to the programming of events. Due to the differences in the frequency of decisions, the scenario stimulus ran for 20 minutes, 30 minutes, or 45 minutes for the high, moderate and low TL conditions, respectively (see Table 2 and Table 4 For a full decision log see Appendix 8a).

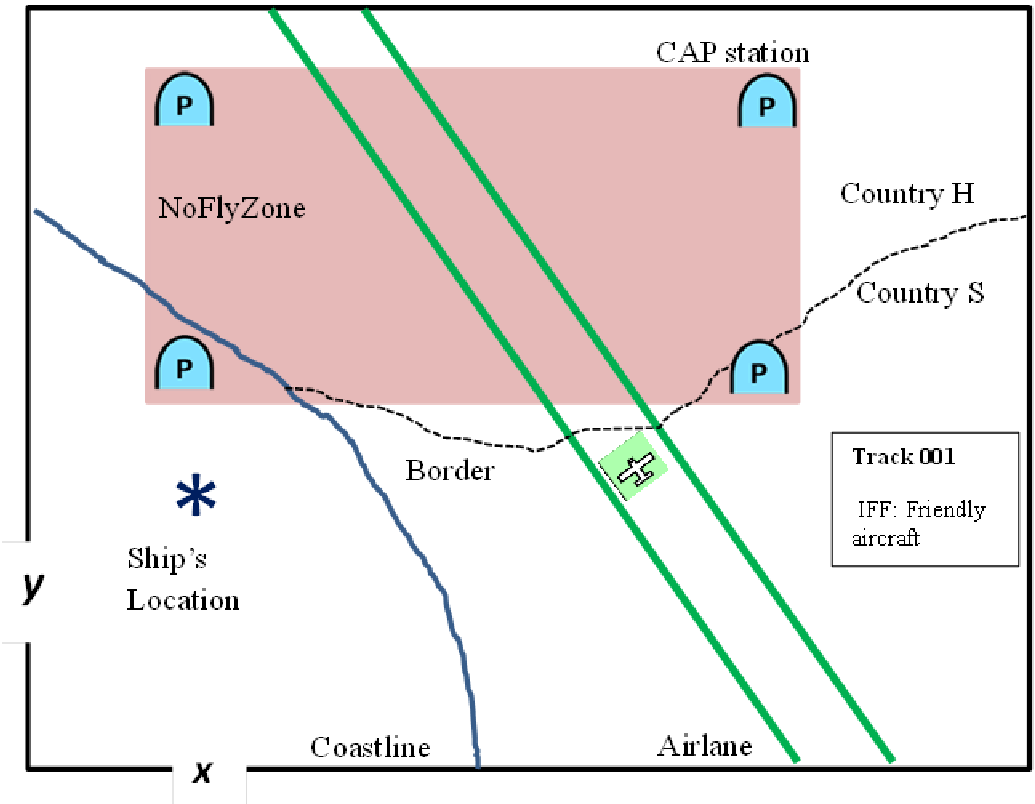


Figure 2. Low TL - 7 minutes into the task

Low TL Condition: In the low TL, all tracks complete their course across the radar screen before the next track appears. Track 001 followed by Track 002, followed by Track 003. In the low TL condition participants only had to monitor one track at a time. Consequently, there was a reduced frequency of the decisions and there were longer periods of inaction where no decision events occurred.

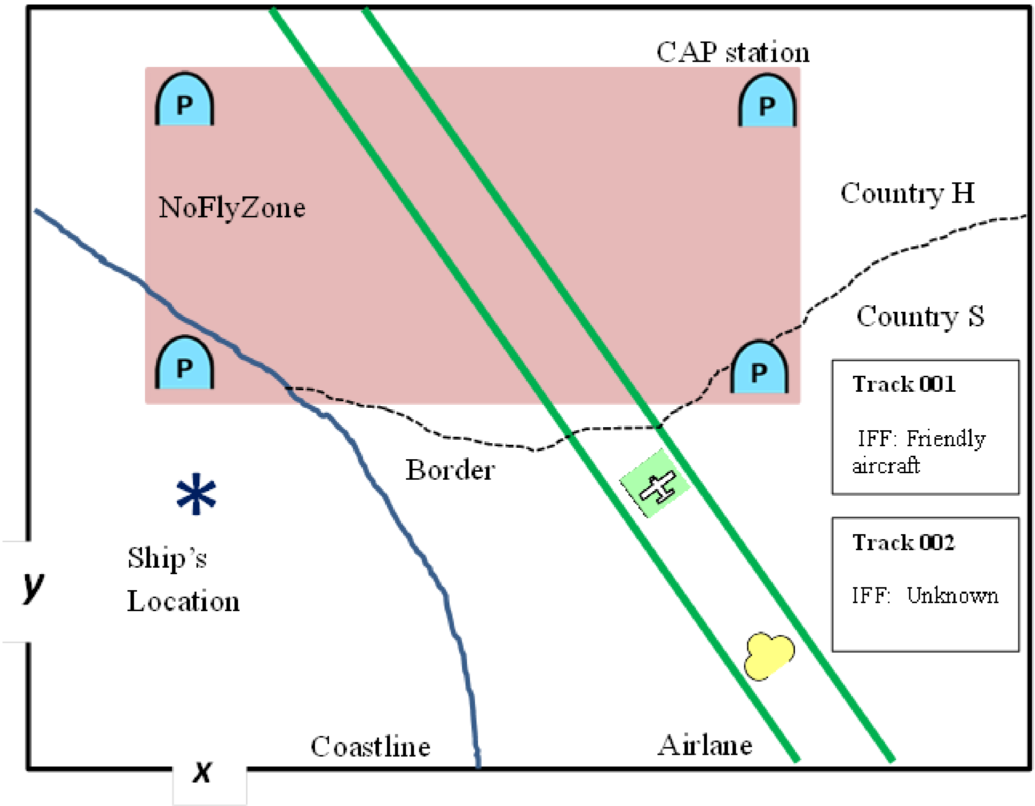


Figure 3. Moderate TL - 7 minutes into the task

Moderate TL Condition: In the moderate TL condition, Track 001 started alone with Track 002 appearing as 001 commences its transit over the NFZ. Track 003 then appears as Track 002 enters the NFZ. In this condition, participants were required to monitor more than one Track at a time and there was an increased frequency to the presentation of the decisions events.

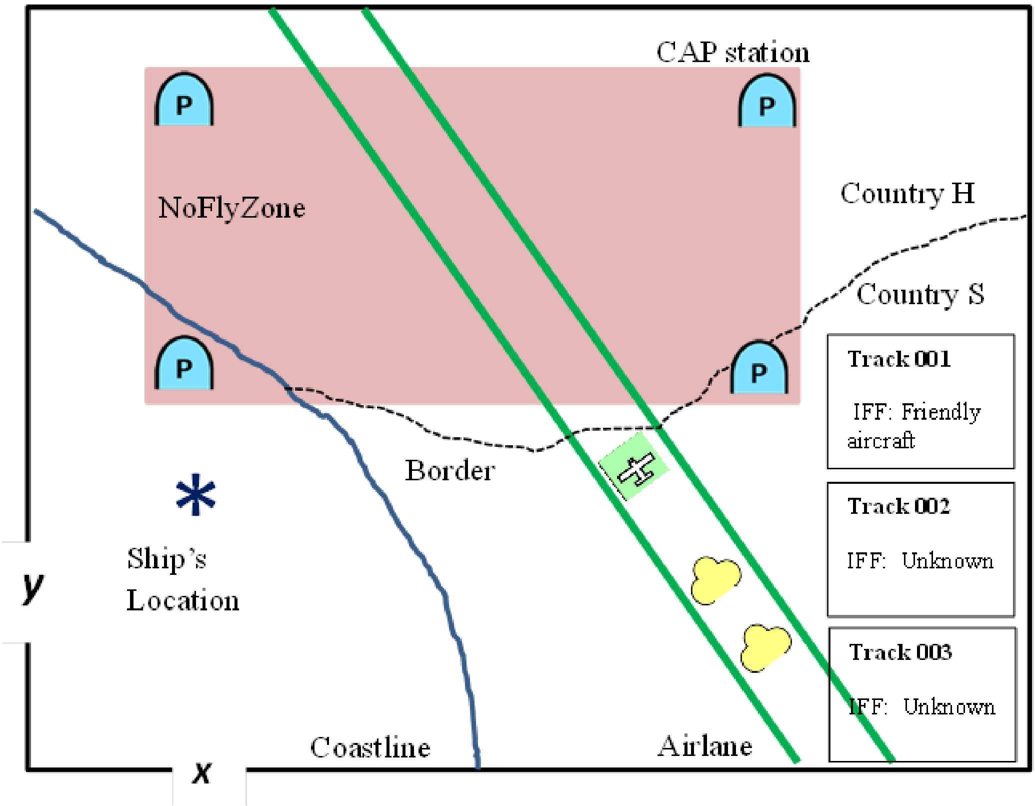


Figure 4. High TL- 7 minutes into the task

High TL Condition: In the high TL condition, all three Tracks appeared simultaneously. In this condition, participants had to monitor and make decisions on all three Tracks. Decision events occurred at a higher frequency and there were reduced time periods of inaction in this condition.

Table 2: Timimgs for the first 10 decision events for each TL condition and correspnding track.

	HIGH TL		MOD TL		LOW TL	
TIME (mins)	**Event**	**Track**	**Event**	**Track**	**Event**	**Track**
0.-1.	1	001	1	001	1	001
1.-2.	2 & 3	001	2 & 3	001	2 & 3	001
2.-3.						
3.-4.	4 & 5	001 & 002	4	001	4	001
4.-5.	6	002				
5.-6.	7	003				
6.-7.	8 & 9	002 & 001	5 & 6	001 & 002		
7.-8.	10	002	7	002	5	001
8.-9.			8 & 9	001 & 002		
9.-10.			10	001		
10.-11.					6	001
11.-12.					7	002
12.-13.					8	002
13.-14.					9	002
14.-15.					10	002

Table 2 displays the timings for the first 10 decision events and corresponding Track it relates to in each TL condition. As displayed in the table, in the high TL, 10 decision events occur in in 8 mins in comparison to 15 minutes in the low TL.

4.2.2.2. DC: Thirty decision events were presented during the experimental simulation. A decision event was defined as; an occasion where a decision needs to be made by an operator. For example, an unknown data-link track (which is the presentation of information received from the data sources of the ship on the radar) appears on the screen. As well as TL, Decision Criticality (DC) was varied across the decision events presented (i.e., 10 high, 10 medium, and 10 low; DC). A decision event with a higher consequence if made incorrectly held a high DC level. A high DC event, for example, could involve an aircraft demonstrating hostile intent, which would have a greater risk if an incorrect decision was made. In comparison, a low DC event, which could involve in identifying a new track on a radar screen, would have less risk associated with it if an incorrect decision was made (see Table 3).

Table 3: Examples of Decision Criticality

Decision Criticality	**Decision Event Example**	**Explanation**
Low	Civil aircraft track continues to adhere to airlane and starts to transit over the NFZ	Identified as civil and following correct procedures. No associated risk.
Medium	Unidentified track reported at 15, 000 ft	Unknown, and flying at an unusual altitude. Medium risk
High	Fighter Ground Attack aircraft continues to close the maritime task group and demonstrates hostile intent.	Hostile aircraft which could cause loss of life. High risk

Table 4: Example of the decision log with decision options

Event Number	Track ID	Event	Decision Criticality	Decision Option 1 BEST DECISION	Decision Option2	Decision Option 3
1	001	A new data link appears on the screen IVO starting position with information consistent with civil aircraft following airlane -300 NM 36,000 altitude	Low	Examine link track data	Validate track is following airlane	Conduct Air "investigate" procedure.

4.2.2.3. Task Instructions and Briefing Materials: Briefing materials were also developed with the SMEs to support the scenarios. These included a mission briefing document, task booklet and task instructions (see Appendices 3c, 3d). The materials provided background information to the participants to assist them in their engagement with the task. The task booklet was developed to provide lay persons with required knowledge related to the processes and terminology employed within the task. The task booklet was specifically created to provide lay persons with knowledge of the tasks that an air defence operator would be familiar with to assist them with their decision making. The booklet included explanations of air defence terminology and radar symbology. A mission brief provided context to the task and task instructions informed participants about the task procedure. These were developed and provided to participants as part of the study's requirements.

4.2.3. Phase Three: Computer Generated Scenarios

Once the scenarios and decision logs had been drafted, a final version was agreed with the SMEs. Following this, the visual display screen was created. As noted, this phase also required familiarization with the software and knowledge needed to produce the scenario videos. The visual display was created using MATLAB®/Simulink and VAPS XT.

4.2.3.1. MATLAB®/Simulink: MATLAB® is a computing programme which allowed the aircraft track trajectories to be prototyped. Computer codes were generated to mimic the flight paths needed for the aircraft. Simulink was then used to read in tracks from MATLAB® and play them into VAPS XT via an appropriate communication protocol.

4.2.3.2. VAPS XT: VAPS XT is a PC-based software tool which was used to generate dynamic, interactive and real-time graphical Human Machine Interfaces (HMI) (Presagis, 2013). For the purpose of the scenarios generated, VAPS XT was used to generate and animate the radar screen display. This involved designing and creating the radar screen, recreating the symbols and connecting the visual display to the correct communications in MATLAB® and Simulink to ensure the appropriate symbol movement.

4.2.3.3. Visual Display: The first step in generating the visual display was to design the overall layout of the scenario with the help of the SMEs, as previously demonstrated in Figure 1. This image included all of the items that needed to be generated. The second step involved prototyping the image creation in MATLAB®. This was completed by Dr Michael Jump. The screen size and items were mapped to where they would be displayed on the screen (see Figure 5). The display was designed for a monitor screen of 1920 x 1080 pixels. VAPS XT was used to create the visual display. The symbols, corresponding track information and the addition of a textbox were all drawn in VAPS XT. Each moving object was associated with an *x* and *y* coordinate and a heading which would be coded to match with an *x* and *y* coordinate in MATLAB®. The *x* and *y* coordinates allowed the positioning of the item and corresponded to the distance in pixels along the East and North axes positions on the radar screen.

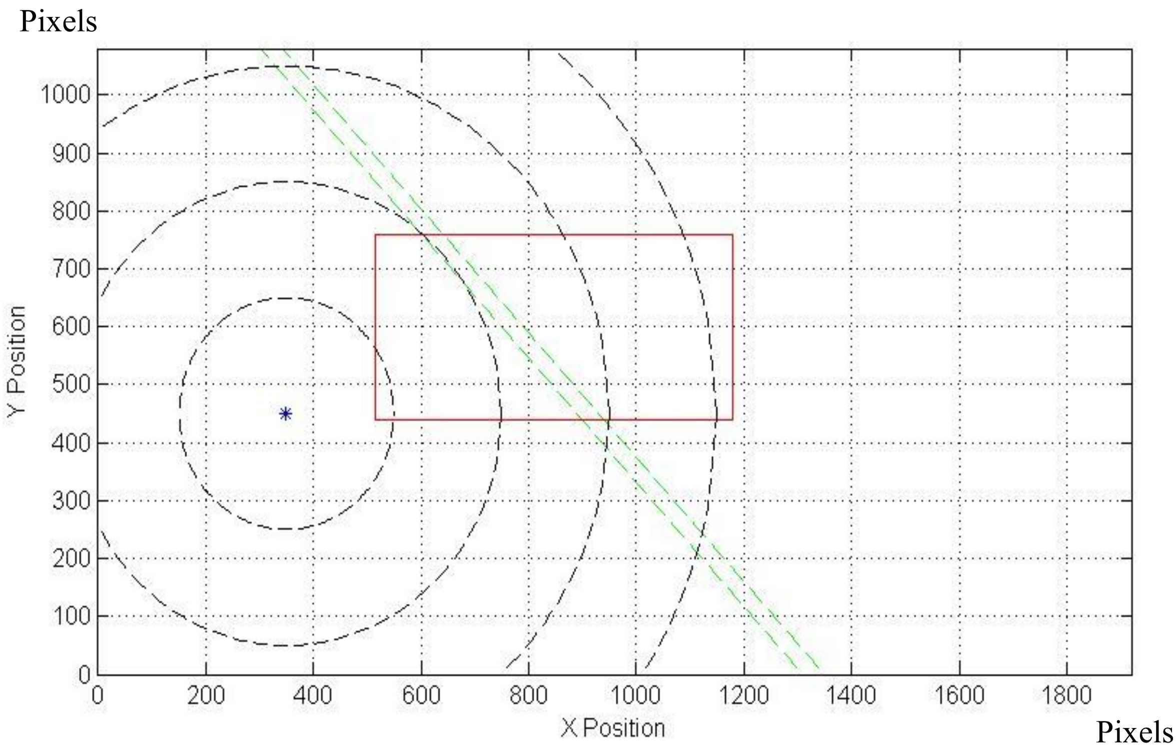


Figure 5. A MATLAB® representation of the radar screen.

The figure displays the design of the radar screen in pixels. A screen size of 1080 x 1920 pixels was used.

4.2.3.4. Aircraft Tracks: To ensure that the tracks moved in a realistic manner, it was necessary to calculate the speed at which the tracks would move across the screen. The trajectory of the aircraft tracks was then coded in MATLAB® to generate *x* and *y* coordinates as well as a heading. A code was written by Dr Michael Jump, which specified the necessary flight profile for Track 001. Due to the complexity of Track 002 and Track 003, these were flown by a student pilot using a flight simulator available in The University's School of Engineering. The resulting *x* and *y* coordinates and headings were then exported to MATLAB® (see Figure 6).

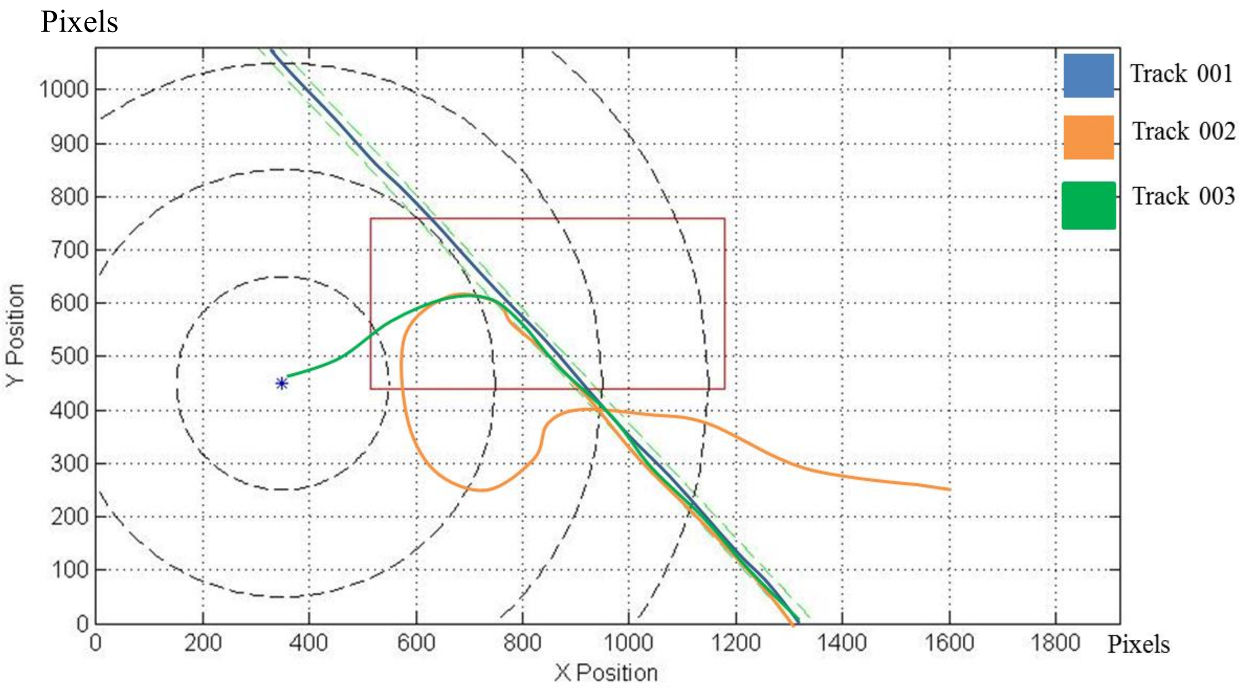
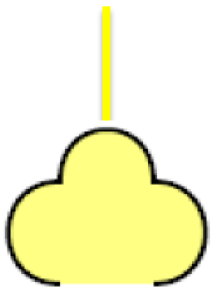


Figure 6. The track trajectories based on *x* and *y* co ordinates generated in MATLAB®

On the visual display, aircraft tracks had Information Friendly Foe, Flight level (IFF), speed and track number attached to them as well as a velocity leader (see Figure 7). The velocity leader was added at the request of the SMEs to increase the fidelity of the visual display (see section 4.3.9). It gives an indication to the direction and speed of an aircraft track. The length of the velocity leader corresponds to the speed of the aircraft; a longer line indicates a faster speed. However, for simplicity, the line was kept at a constant length in the scenarios used.



TR: Track number

IFF: Unknown, Friendly or Hostile

FL: Flight level e.g. FL150 = 15 000ft

SP: Speed in knots

Figure 7. VAPS generated unknown data track symbol with attached information which was displayed on the screen.

4.2.3.5. Symbology: The symbology used in the study is as specified by APP-6c (North Atlantic Treaty Organisation [NATO], 2008). The symbols were then recreated in VAPS XT to be representative of the correct symbols. Only the symbols needed in the task were created and shown to the participants. The main symbols used were, a) unknown data track, b) hostile aircraft [Fighter Ground Attack (FGA)], c) civil aircraft (CIVAIR) and d) Combat Air Patrol (CAP) stations (see Figure 8).

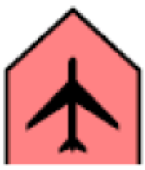
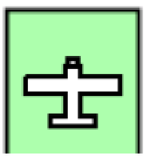
	UNKNOWN DATA LINK/TRACK	**HOSTILE AIRCRAFT/FGA**	**Civil Aircraft/ CIVAIR**	**CAP stations**
				

Figure 8. NATO Symbology

4.2.3.6. Scenario Video: Once the visual display had been created in VAPS XT and linked to the data in MATLAB®/Simulink, the next step was to create the experimental video that could be played to the participants. Live Gamer HD, which is a video capture card, was used to record and then generate a video file. Microsoft Movie Maker was used to edit the video to include the pauses required by the experimental plan. The pauses allowed the participants to locate the correct question in the questionnaire for a given decision event. It has previously been shown that a pause does not cause an interference with the task (Endsley, 1995). In addition, on-screen timers were created to indicate to the participants how much time left they had to make a decision.

4.2.3.7. Practice Trial

Before taking part in the experimental scenario, participants undertook a practice trial, created using the same steps as described above. The practice trial was a short scenario which consisted of 5 decision events which lasted 5 minutes. The decision events in the practice trial allowed the participant to familiarise themselves with the task and the procedure. For example, one decision event required the participant to locate the ship's position on the screen. The duration of the trial was limited so as not to fatigue the participants before the experimental task (Barnes-Yallowley; personal communication, 2015).

Before discussing Phase Four, which involved the conduct of a pilot study, the design and measurements used throughout the Experimental Work are presented.

4.3. Experimental Methods

4.3.1. Participants: In total, 322 participants were recruited through opportunity sampling. Three hundred were novice participants. These had no prior experience of naval air warfare. Twenty-two were RN PWO personnel who had recently completed a PWO training course and had, on average, 9 years' experience at sea. Novice participants were invited to get in touch with the researcher via emails, University announcements and advertisements. The sample consisted predominately of psychology undergraduates. However, other students at the University of Liverpool who responded to emails, announcements and advertisements were also recruited. The novice group were categorised into three groups: Non-Gamers; Video Game Players 1 and Video Game Players 2. The groupings are described in more detail, below, in section 4.4.3.2. VGP populations were invited to get in touch with the researcher via emails, university announcements and advertisements, as well as social media campaigns specifically advertising for participants who play video games.

Twenty-two RN personnel were recruited via a gatekeeper at Dstl. Participants were recruited towards the end of PWO courses run at the Maritime Warfare School (MWS). Defence Council Instructions (DCI) and Temporary Memoranda were used to target the correct level of decision maker. This was facilitated by Dstl's Military Advisors assigned to the project. Personnel interested in hearing more about the study were sent the Participant Information Sheet (PIS) for more information, which included the purpose of the study, what they would be asked to do (see Appendices 3a & 3b). If they remained interested, they were invited to contact the investigator to make an appointment to sign the consent form and participate in the study. They were given a period of at least 24 hours to consider the PIS before being invited to sign the consent form.

4.3.2. Ethics: Ethical approval was obtained from The University of Liverpool (see appendix 1a). In addition, to comply with the MoD's ethics policy, firstly a Scientific Assessment Committee (SAC) was attended. The SAC is to ensure that scientific and technical rigour is assured. Following a favourable opinion from the SAC, a full Ministry of Defence Research Ethics Committee (MoDREC) was attended. The committee is an independent body comprised of non-MoD (expert and lay) members. A favourable opinion was received from MoDREC in February 2017 (see Appendix 1b). This allowed the research experiments to go ahead as they met the ethical standards set out by the MoD.

4.3.3. Independent Variables:

4.3.3.1. TL and DC: As previously mentioned, TL in the experiments referred to the cognitive and temporal load of the condition. In the task, this was manipulated by varying the number of tracks displayed on the radar screen and the frequency and speed at which decision events occurred. The high TL condition involved an increased frequency of decisions, multiple decisions and more aircraft tracks to monitor on the screen. In comparison to the low condition which was characterised by only one aircraft track to monitor at a time, decisions based only on the one track, reduced frequency of decision events and longer periods of inaction. The criticality of a decision was categorised on the basis of the associated level of risk of an incorrect decision For example, a decision event with a higher consequence if made incorrectly held a high DC level.

4.3.3.2. Group: A set of demographic questions asked all participants about their sex, age, whether they played video games and how many hours on average they played a week (see appendix 5a). For experiments 1, 2 and 3, based on this information, participants were then divided into 3 groups, Non-Gamers (NG), Video Gamer Players 1 (VGP1) and Video Gamer Players 2 (VGP2). NG consisted of individuals who reported that they did not play any video games on a regular basis. The second group, VGP1, consisted of individuals who played up to 7 hours a week VGP2 were gamers who played on average more than 7 hours a week and reported as having done so consistently over the last 2 years. This definition of VGPs was taken from Boot et al. (2008). Military personnel consisted of PWOs who had attended and completed a PWO training course at HMS Collingwood. For these participants, additional information was collected on years in service, rank, previous roles and years at sea (see Appendix 5b).

4.3.3.3. Time Pressure: Time pressure refers to the time allocated to individuals to make a decision. This was a countdown from either 20 or 10 seconds, indicated by a timer on the screen.

4.3.3.4. Audio: Participants were randomly allocated to either an 'audio present and attending' or 'audio present as background noise' condition. Participants in each condition were instructed prior to the task as to whether they were required to attend to the audio or not. The audio sample involved typical communications between a pilot and Air Traffic Control (ATC) during a general aviation flight using Very High Frequency (VHF) radio communications. The audio, therefore, included a conversation between pilot and passenger, pilot and ATC and ATC and other aircraft on frequency. There are breaks in conversation, overlapping conversations, interruptions and the background noise typical of VHF communications, etc. The audio was not related to the task but is similar to the background noise that an operator would be subjected to on the communication networks accessible to a ship's Ops room. The audio was played throughout the duration of each scenario. Naturally, due to the nature of the audio sample, some decisions were being made with audio present and others not. This was dependant on the running order of the task scenario. The audio was the same for all TL conditions. At the end of the scenario, all participants (Attend and Non Attend) were asked questions related to the audio that was played to them (see Appendix 8d). This was to assess the task manipulation of attendance.

4.3.3.5. Feedback: Individuals in the Metacognitive Feedback Training (MFT) condition received a PowerPoint-briefing prior to the task (see Appendix 8e) and, after the practice trial participants received feedback on their performance. The information in the PowerPoint-briefing included information which explained what was meant by the C-A relationship. It also highlighted the importance of ensuring that the confidence placed in a decision was directly related to the accuracy of that decision. An example from a forensic setting was provided to help demonstrate the importance of this relationship. This was an example of being a witness to a crime and having to give evidence. Following the PowerPoint briefing, the participants then took part in a practice trial. After the practice trial had been completed, the researcher then provided verbal feedback on the participant's responses to indicate whether they had the correct levels of confidence in their decisions.

4.3.4. Dependent Variables:

4.3.4.1. Confidence: Participants were asked to rate confidence in their decision on a Likert scale, '0' being *not confident at all* and '5' being *extremely confident*. As there were 30 decision events, the maximum confidence score a participant could record in the task was 150. The 30 events varied in DC (high, medium, low). As there were 10 decision events for each DC, a maximum score of 50 could be recorded.

4.3.4.2. Accuracy: The accuracy of the decisions was decided on by the SMEs. When designing the decision log and generating the options, one of the options was decided to be the best decision, given the current situation. Participants were scored '1' for an *accurate response* or '0' for an *incorrect response*. As there were 30 decision events, the maximum total possible was 30. As there were 10 decision events for each DC, a maximum score for each DC was 10.

4.3.4.3. W-S C-A: The answer to each question was coded as correct or incorrect. The confidence score for each question was recorded in order to generate a numerical relationship between confidence and accuracy for each participant (i.e. a point-biserial correlation).

4.3.4.4. Psychometrics, Workload, Situational Awareness: The next section will describe these measures in more detail.

4.3.5. Materials

4.3.5.1 Peace Enforcement Scenario: This is described in Section 4.2.1.1.

4.3.5.2. Task Instructions and Briefings: This is described in Section 4.2.2.3.

4.4.6. Personality and cognitive constructs and performance measurements

Several questionnaires and standardised measures were used. This section will describe the main questionnaires used throughout the experiments in this thesis. The aim was to uncover the potential related individual differences in operators and whether any of these measures relate to decision confidence, accuracy and metacognitive abilities. The author was trained in the administration and interpretation of the tools by an authorised user who was also one of the author's supervisors.

4.3.6.1. Personality: To assess personality constructs, the revised NEO personality inventory (NEO-PI-R) (Costa & McCrae, 1992) was used. The NEO-PI-R is a five-factor model of personality and consists of the major personality factors. It measures five broad personality constructs 1) Neuroticism, 2) Extraversion, 3) Openness to experience, 4) Agreeableness and 5) Conscientiousness. All items are scored 0-4 where *Strongly Agree* = 0 and *Strongly Disagree* = 4. A number of items are reverse scored. The NEO-PI-R has strong support for reliability, construct and discriminate validity (see Costa & McCrae, 1992; Piedmont & Weinstein, 1993). Internal consistency coefficients are reported as 0.86 to 0.95 for both self and observer. Furthermore, neuroticism, extraversion and openness to experience had good long-term test-retest reliability. All five had high short-term test-retest reliability (Costa & McCrae, 1992).

4.3.6.2. Cognitive Constructs: A Tolerance of Ambiguity questionnaire was used to assess individual Tolerance of Ambiguity (Budner, 1961). This scale measures how comfortable respondents are with ambiguity. There are 16 items rated on a Likert scale where 0 = *Strongly Agree* and 4 = *Strongly Disagree*. The items included an equal number of positively and negatively worded items. All scores are added with a number of items being reverse scored. The average score was 44-48; higher scores indicate a greater intolerance to ambiguity (Budner, 1961). This measure has been found to be free from acquiescent and social desirability response tendencies and validated with a good test correlation.

A questionnaire for Decision Style, separated into elements of tolerance (high scores = *less tolerant*) and decisiveness (high score = *more* *decisive*) were taken from the Need for Closure questionnaire (Roets & Van Hiel, 2007). Tolerance of Ambiguity and Decisiveness combine to provide a scale for a decision style which explicitly probes the need for quick and unambiguous answers.)

4.3.6.3. WL: NASA Task Load index [TLX; see Appendix 2b] (Hart & Staveland, 1988) is a subjective measure where the experiment participant rates their perceived workload during a task. NASA TLX has been used in a wide variety of tasks and contexts and has good reliability and validity. NASA TLX was undertaken upon completion of the scenario. Firstly, participants were asked to complete a pairwise comparison of the six (6) dimensions: 1) Mental Demand, 2) Physical Demand, 3) Temporal Demand, 4) Performance, 5) Effort, and 6) Frustration (Table 5). For each pair, participants were asked to choose the dimension which represented the most important contributor to workload for the task. For example, choose between Effort or Performance. For this, there were 15 possible comparisons. The number of times a scale was chosen was tallied, providing a weighting for that particular dimension. Secondly, participants were asked to rate the 6 dimensions of workload on a 20 point scale. The increments increased by a factor of 5 (0 = *Low* – 100 = *High*). This provided a rating for each dimension. The overall workload score is then derived by multiplying each rating by the weight given. The sum of the weighted ratings is then divided by 15.

Table 5: Descriptions of the subscales of NASA TLX

Scale Title	Description
Mental demand	How much mental and perceptual activity requires (e.g. thinking, deciding, calculating remembering, looking, searching, ect.).
Physical demand	How much physical activity was required (pushing, pulling, turning, controlling, activating, etc)? Was the task easy or strenuous, restful or laborious?
Temporal demand	How much time pressure did you feel due to the rate or pace at which the tasks or Trask elements occurred? Was the pace slow and

	leisurely or rapid and frantic?
Performance	How successful do you think you were in the accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals.
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	How insecure, discourages, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

4.3.6.4. SA: SA was measured by Situation Awareness Rating Technique [SART] (Taylor, 1990). The SART was administered upon completion of the scenario. It involves the participant rating each of the 10 dimensions (see Table 6) on a seven-point rating scale (1 = *Low*, and 7 = *High*). The ratings were then combined in order to calculate a measure of participant SA. An overall SART score is calculated using the following formula: SA = U – (D – S), where: U = *summed understanding*; D = *summed demand*; S = *summed supply*. SART has shown to be sensitive to task difficulty (Selcon et al., 1990) although there has been a lack of test – retest reliability (Vidulich, Crabtree & McCoy, 1993). Nevertheless, SART is advantageous as it is easy to administer, can be carried out on a range of tasks and is good for predicting performance (Endsley, 1998).

Table 6: Descriptions of the SART dimensions

Scale	**Subscales**	**Description**
Demand	Instability of Situation	Likeliness of situation to change suddenly
	Variability of Situation	Number of variables which require your attention
	Complexity of Situation	Degree of complication(number of closely connected parts) of the situation

Supply	Arousal	Degree to which you are ready for activity
	Spare Mental Capacity	Amount of mental ability available to apply to new tasks
	Concentration	Degree to which your thoughts are brought to bear on the situation
	Division of Attention	Amount of division of your attention in the situation
Understanding	Information Quantity	Amount of knowledge received and understood
	Information Quality	Degree of goodness or value of knowledge communicated
	Familiarity	Degree of acquaintance with the situation
	Instability of Situation	Likeliness of situation to change suddenly

4.3.7. Procedure: This section will describe the general procedure for experiments 1, 2 and 3. Before taking part in the experiment, all participants were given time to read the information sheet. Once they were happy to take part, they completed a consent form. Participants were randomly allocated to a high, moderate, or low TL condition. Participants first completed participant demographic forms which collected data on age, gender and occupation. Participants were also asked to complete paper-based questionnaires to gauge the relevance of a number of measures across groups (e.g., general personality constructs, thinking and reasoning) where they may be relevant to particular questions. Following this, participants were provided with the task booklet to read. The task booklet provided the participants with the information needed to assist them in the decision making task, including air defence terminology and symbols. Once they had read the booklet, participants undertook a practice trial. The practice trial involved a series of decision events which allowed the participant to familiarise themselves with the task and the procedure. The questionnaire booklet presented three separate decision options based on the events of the scenario. One choice was required to be selected by placing a tick by the option they believed to be the ‘best option given the current situation’.

Participants were then required to rate how confident they were in the options chosen on a Likert scale, where 0 = *not at all confident* to 5 = *extremely confident*. After either 10 or 20 seconds, depending upon the time manipulation, the screen was blanked out to signal to the participants that the allocated decision time had ended. All participants then undertook the experimental air defence scenario, following the same procedure as described for the practice.

Thirty decision events were presented during the experimental simulation. As previously described above, DC was varied across the decision events (i.e., 10 high, 10 medium, and 10 low; DC). The scenario stimulus ran for 20 minutes, 30 minutes, or 45 minutes for the high, moderate and low conditions, respectively. Once the scenario had finished, participants completed the Situational Awareness and Workload questionnaires. Participants were fully debriefed to ensure each understood the nature of the study and given the opportunity to ask further questions.

4.3.8. Phase Four: Pilot Study: A pilot study was conducted to act as a shake-down trial to check the experimental set-up. This was done to ensure that the expected differences in the TL conditions were met and that the task could be carried out by the participants. The results from the pilot study are reported below.

4.3.8.1. Participants: In total 11 participants were recruited for the pilot study, 5 females and 6 males with a mean age = 21 (*SD* = 4.68). All participants were students from the University of Liverpool. As noted earlier, the research was approved by both University of Liverpool and MoD ethics committees. Opportunity sampling was employed. Participants responses were kept confidential and were only identified by a number on their consent form and answer sheets.

4.3.8.2. Results: The results highlighted the success of the experimental set up and procedures. The TL conditions produced significant differences in levels of WL. In the high and moderate TL workload was found to be higher and SA was lower than in the low TL conditions. The results are firstly discussed in relation to WL and SA, to assess the task manipulations.

The results are also displayed for confidence, accuracy and W-S C-A to examine the general trends in data are also shown.

4.3.8.3 WL and SA: A one-way analysis of variance (ANOVA) was conducted to assess the differences in WL across the TL conditions. A significant difference was found in TL $F(2,8) = 10.13$ $p = .006$. The results demonstrate higher WL in the high TL ($M = 59.50$, $SD = 4.93$) than the low TL ($M = 29.25$, $SD = 14.84$), $p = .021$ as well as higher WL in the moderate TL ($M = 66.00$, $SD = 14.00$) compared to the low TL, $p = .011$. WL was found to be higher in the moderate TL condition and lowest in the low TL condition. The results suggest that high and moderate TL conditions were more demanding than low TL.

Additionally, although no significant differences were found in SA $F(2,8) = 1.43$, $p = .295$, individuals reported the highest SA in the low TL condition ($M = 19.25$, $SD = 4.78$) and the lowest report was found in the moderate TL ($M = 13.67$, $SD = 5.03$), suggesting individuals felt they had better SA in the low TL condition.

Table 7: Means and Standard Deviations for WL and SA as influenced by TL

TL	Overall WL	Overall SA
High	59.50 (4.93)	15.25 (4.03)
Moderate	66.00 (14.00)	13.67 (5.03)
Low	29.25 (14.84)	19.25 (4.78)
Total	50.27 (19.94)	16.27 (4.77)

Note. Standard Deviations are in parenthesis.

4.3.8.4. NASA TLX subscales: The subscales were analysed to examine any differences in TL conditions on the 6 subscales. The means and standard deviations are shown in Table 8. Descriptive statistics show that participants rated the task to be mentally demanding in all TL conditions. Examining the TL conditions, the low TL condition was lowest in mental demand, temporal demand, effort, and frustration. In the high TL condition, individuals reported high results for frustration and effort. In comparison, the moderate TL condition produced the highest levels of WL for mental demand, temporal demand and performance. These results suggest that the different TL conditions impose different demands on individuals. Although these results suggest that the moderate TL was higher in demand, due to the small sample size and closeness between high and moderate TL, these categories were kept the same.

Table 8: Means and Standard Deviations for dimensions of WL as influenced by TL.

TL	Mental Demand	Temporal Demand	Effort	Physical Demand	Performance	Frustration
High	313.75 (50.23)	148.75 (93.66)	175.00 (59.116)	0.00 (0.00)	105.00 (50.50)	150.00 (54.77)
Moderate	391.67 (38.19)	188.33 (156.07)	170.00 (132.29)	0.00 (0.00)	136.67 (57.52)	108.33 (48.56)
Low	107.50 (115.87)	87.50 (87.03)	67.37 (61.17)	3.75 (7.50)	121.25 (65.37)	55.00 (97.13)
Total	260.00 (144.00)	137.27 (107.71)	134.55 (92.21)	1.36 (4.52)	119.55 (53.69)	104.09 (77.55)

Note. Standard Deviations are in parenthesis.

4.3.8.5. SART subscales: The descriptive statistics of the subscales of SA found that the moderate TL was found to be the most demanding and required the most supply of resources. The TL conditions demonstrated similar levels of understanding (see Table 9).

Table 9: Means and Standard Deviations for dimensions of SA as influenced by TL

TL	Demand	Supply	Understanding
High	14.00 (4.08)	17.75 (4.57)	11.50 (2.52)
Moderate	16.67 (1.53)	19.33 (3.22)	11.00 (1.73)
Low	9.75 (4.03)	17.50 (4.04)	11.50 (1.73)
Total	13.18 (4.35)	18.09 (3.73)	11.36 (1.86)

Note. Standard Deviations are in parenthesis

4.3.8.6. Accuracy: A 3 X 3 ANOVA was conducted on the accuracy data. A main effect of TL was found F(2,8) = 7.58, p = .014, η_p^2 = .66. Participants were more accurate in the low TL condition (M = 5.58, SD = .85) compared to high (M = 4.17, SD = .31) p = .007. Also, they were more accurate in moderate TL condition (M = 5.44, SD = .41) than high, p = .017. No main effect of DC F(2,16) = .69, p = .515, η_p^2 = .08 was found, and no interaction was observed, F(4,16) = 1.38, p = .286, η_p^2 = .26 (see Table 10).

Table 10: Means and Standards Deviations for Accuracy as influenced by DC and TL

TL	Overall Accuracy	High DC	Medium DC	Low DC	Total
High	12.50 (1.29)	4.50 (.58)	4.25 (.96)	3.75 (1.50)	4.17 (.31)
Moderate	15.33 (1.53)	5.00 (1.73)	5.33 (.57)	6.00 (1.73)	5.44 (.41)
Low	16.75 (1.50)	6.75 (.96)	5.25 (.96)	4.75 (.96)	5.58 (.85)
Total	14.82 (2.32)	5.45 (1.44)	4.91 (.94)	4.73 (1.55)	5.06 (.28)

Note. Standard Deviations are in parenthesis.

4.3.8.7. Confidence: No significant differences were found in the confidence data in DC $F(2, 16) = 1.93$, $p = .178$, $\eta_p^2 = .194$. No interaction was found between TL and DC $F(4, 16) = 2.29$, $p = .104$, $\eta_p^2 = .36$. No main effect of TL $F(2, 8) = 2.13$, $p = .181$, $\eta_p^2 = .348$. However, overall participants were more confident in the low TL condition (*M* = 41.17, *SD* = 3.41). Participants were equally confident in the high (*M* = 32.75, *SD* = 1.15) and moderate (*M* = 32.89, *SD* = 2.14) TL conditions. The data suggests individuals display higher levels of confidence in the low TL, and specifically in low DC decisions. Similar confidence ratings are demonstrated across DC levels.

Table 11: Means and Standard Deviations for Confidence as influenced by DC and TL

TL	Overall	High DC	Medium DC	Low DC	Total
High	98.25 (10.05)	31.50 (3.11)	33.00 (4.83)	33.75 (3.30)	32.75 (.94)
Moderate	98.767 (34.93)	35.33 (7.02)	32.00 (12.12)	31.33 (15.82)	32.89 (1.75)
Low	123.50 (9.26)	42.75 (25.32)	37.25 (3.78)	43.50 (1.29)	41.17 (2.79)
Total	107.55 (21.45)	36.64 (6.87)	34.27 (6.81)	36.64 (9.19)	35.60 (3.94)

Note. Standard Deviations are in parenthesis.

4.4.8.8. W-S C-A. To establish if there were any significant effects of TL (High, Moderate, Low) and level of Decision Criticality [DC] (Low, Moderate and High) for W-S C-A correlations, it was first necessary to calculate each individual participant's W-S C-A score. The responses provided by participants to each question were coded as correct or incorrect, and the confidence score for each question was recorded to generate a numerical relationship between confidence and accuracy. Only the TL was analysed for the pilot study. No significant differences were found in between the TL conditions $F(2, 8) = .12$, $p = .889$.

Table 12:Means and Standard Deviations for W-S C-A as influenced by TL

TL	Overall W-S C-A
High	.07 (.18)
Mod	.01 (.21)
Low	.06 (.10)
Total	.05 (.15)

Note. Standard Deviations are in parenthesis.

4.3.8.9. Discussion: The results from this small pilot study demonstrate a successful manipulation between the different conditions and that participants were able to complete the experiment. The results demonstrated that individuals were more accurate in the low task condition than its high counterpart. This finding could suggest that the lower TL allowed the individuals to make more accurate decisions as there was less cognitive load on them. However, no main effect of DC was observed in the pilot study. DC did not impact on decision accuracy, confidence or metacognitive ability. Furthermore, the data suggests individuals display higher levels of confidence in the low TL condition, and specifically in low DC decisions. Again, this suggests that a reduced cognitive and temporal load on the individual allowed the individuals to make decisions more confidently. To examine individual awareness of their accuracy W-S C-A was computed. The findings from W-S C-A was relatively low, and descriptive statistics show that both high and low TL produced similar levels of W-S C-A. Individuals performed worse in the moderate TL condition. In summary, the pilot study demonstrated that the task could be completed and understood by participants, as well as ensuring that there were varying degrees of WL and SA in the different TL conditions. The low TL was associated with lower WL and higher SA.

4.3.9. Phase Five: Completed SME Validated Scenarios and Decision Logs: A final meeting with the SMEs was held to finalise the necessary materials for the study and to ensure that the experimental set-up was correct. Due to the availability of the SMEs, this was conducted after the pilot study. The meeting led to some minor adjustments to the visual display. This included the addition of a timer, to give participants an indication of how much time they had to make a decision and to function as a method to increase pressure in the appropriate conditions. Additionally, to increase ecological validity, speed vectors and track information were added to the visual display next to the respective tracks (see Figure 7). Once this work had been completed, this concluded the development and the Experimental Work began.

In summary, this Section of the Chapter has discussed the steps involved in generating the experimental scenario, detailing each Phase. In addition, the results of the pilot study were also included. The completion of Phases 1-5 took approximately one year, with time being spent familiarising with and understanding the doctrine and procedures in air defence decision making as well as the acquisition of knowledge of the software and skills required to build the computer simulations for experimental testing.

Chapter 4 has discussed the steps involved in the development of the experimental stimulus and illustrated the various Phases involved. This section concludes Part 1 of the thesis. Part 2 of the thesis is the Experimental Work. Chapters 5-10 will present the experiments conducted and which address the aims of the thesis.

PART TWO: The Experimental Work

Introduction

Part Two of the thesis contains the Experimental Work which is divided into two sections. Section A: Foundation and Investigatory Work (Chapters 5-8) and Section B: Application (Chapters 9-10). Section A of the Experimental Work established the potential factors that may relate to decision accuracy, confidence and W-S C-A. The first experiment (Chapter 5) provides the baseline for the Experimental Work. This experiment aimed to establish the influence of TL, DC and individual differences including group, personality and cognitive constructs in the air defence context. Following this, and to introduce an element of increased time pressure, the time to make a decision was reduced for the second experiment (Chapter 6). Next, to examine if there were any differences between these two time conditions (Experiment 1 and Experiment 2), an additional analysis of these data sets was conducted (Chapter 7). The third experiment (Chapter 8) aimed to examine the influence of audio on the variables under interest. As such participants were divided into two groups. Both groups were instructed to wear headphones which played an audio conversation between two pilots, unrelated to the task. One group was instructed to attend to the audio and the other group were instructed not to attend. Following this, section B of the Experimental Work was concerned with the application of this measurement; as such Experiment 4 (see Chapter 9) introduced MFT element prior to the task to assess whether this element had an impact on decision accuracy, confidence and W-S C-A. To conclude, the fifth experiment (Chapter 10) examined the application of this experimental set up to SMEs. As such, the same experimental set up as Experiment 1 was conducted on PWOs from the RN.

SECTION A: Foundation and Investigatory Work

CHAPTER FIVE

5. Investigating the Impact of DC, TL and group on Decision Accuracy, Confidence and W-S C-A

5.1. Introduction

The literature presented in Chapter two discussed the impact of a range of factors that have been shown to influence decision accuracy, confidence and W-S C-A. In light of this, the first experiment aimed to begin to uncover some of these factors in relation to the air defence domain. Specifically, it examined the external factors of DC and TL and internal factors of personality, cognitive constructs and gaming experience together with the implications of these factors on confidence, accuracy and W-S C-A.

DC is potentially a key factor in decision making however, it has not previously been considered in much detail. Research into DC has been inconsistent as findings have shown that criticality can have both a negative and positive impact on performance (Hanson et al., 2014; Callister et al., 1999). Furthermore, of the existing research, the investigation of DC has been considered in relation to a scenario but not the individual decision event. Needless to say, an individual decision event could have a considerable impact on an individual's decision making ability. For instance, the criticality of a decision event was found to influence decision accuracy, confidence and W-S C-A (Adams-White et al., 2018). Hence DC, in an air defence domain, will be considered in this experiment.

Another factor of interest in this research is TL. A review of the literature demonstrated that TL is an important factor to consider that may influence decision making. TL refers to the cognitive and temporal load exerted on the individual and research has shown that this may influence decision making ability, as well as increasing experience of stress and WL on the individual. This current experiment, therefore, aimed to examine the conditions which mediate the relationship between confidence and accuracy and the impact of TL on decision making in the air defence context.

As well as considering the external factors, this experiment was also interested in individual differences in air defence decision making. As previously discussed, individual differences in decision making have been shown to influence both decision accuracy and confidence. As such, psychometric measures have been used to asses this relationship. In particular personality traits as well as cognitive constructs related to uncertainty were investigated. Furthermore, as discussed in Chapter two, Adams-White et al. (2018) investigated both DC and TL and those individual traits discussed above however; other recent research has also shown that VGPs may be suitable for different roles and display different cognitive abilities that may be relevant to the air defence environment (Wheatcroft et al., 2017). Hence, the experiment reported here expands from Adams-White et al. (2018) to examine differences in non-naval populations of students and VGPs.

In addition, it is beneficial to understand performance via a range of different measures as well as assessing their relationship to decision accuracy, confidence and metacognition. Hence, this experiment also aimed to establish how the variables under investigation align with the wider measurements currently used in human-machine interaction decision making literature (WL and SA). SA and WL were also considered to ensure differences between the TL conditions.

5.2. Hypotheses (HP)

- **HP1:** High DC will reduce decision accuracy (A), confidence (B) and W-S C-A (C).
- **HP2**: High TL will reduce decision accuracy (A), confidence (B) and W-S C-A (C).
- **HP3:** Individual differences will be found in decision accuracy, confidence and W-S C-A.

- **HP4:** VGPs will be more accurate (A), confident (B) and display higher W-S C-A(C).
- **HP5:** To explore the relationship between accuracy (A), confidence (B), W-S C-A (C) and measurements of WL and SA (D).

5.3. Participants

Ninety participants were recruited through opportunity sampling from the University of Liverpool. The participants consisted of 33 females and 57 males with a mean age of 24 years (*SD* = 4.42). Using a demographics form, participants were required to state whether they played video games and how many hours on average they played a week. Based on this information participants were then divided into three groups, Non-Gamers (NG), Video Gamer Players 1(VGP1) and Video Gamer Players 2 (VGP2). The NG group consisted of individuals who reported not playing any video games. The second group (i.e. VGP1) consisted of individuals who played video games up to seven hours a week, and VGP2 were gamers who played video games on average more than seven hours a week and reported as having done so consistently over the last two years. This definition of VGPs was taken from Boot et al. (2008). Each group consisted of 30 participants. None of the participants had any prior experience in naval warfare operations as the study was initially interested in the Ops Room role and novice capacity to the task. The sample size was decided upon by design, power and previous studies using G power analysis with an effect size of .8 and significance level of .05 (Faul, Erdfelder, Lang & Buchner, 2007). The study received approval from the University of Liverpool's Institute of Psychology Health and Society Ethics Committee and a favourable opinion from MoDREC.

5.4. Results

To assess the differences in means a number of statistical analyses were performed on the data for accuracy, confidence and W-S C-A using Analysis of Variance (ANOVA). A manipulation check was carried out to assess the differences in TL (see analysis of WL and SA). Significant differences were found between TL as suggested by differences in reported WL, $p < .001$ and SA, $p = .003$. As such we can assume that the TL manipulation was significant. An alpha level of .05 was used for all statistical tests unless stated otherwise.

5.4.1. Accuracy:

The accuracy of the decisions was decided upon by the SMEs (see Chapter 4). When designing the decision log and generating the decision options, one of the decision options was voted as the best decision given the current situation and circumstances. Participants scored '1' for *an accurate response* or '0' for *an incorrect response*. The maximum total was 30 and the maximum mean for each DC was 10. To examine the mean differences between DC, TL and group in accuracy ANOVA was performed on the data.

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data (see Table 13).

A main effect of DC was found on decision accuracy $F(2,162) = 18.84$, $p < .001$, $\eta_p^2 = .19$. Bonferroni corrected post hoc tests showed participants were more accurate in high DC decisions ($M = 5.26$, $SD = 1.89$) than both medium DC ($M = 4.62$, $SD = 1.73$), $p = .036$ and low DC decisions ($M = 3.69$, $SD = 1.89$), $p < .001$. A significant difference also existed between medium and low DC, $p = .001$. The findings demonstrate that decisions that were more critical were made with more accuracy. Hence, the high DC increased decision accuracy and thus, did not support the HP1-A.

A main effect of TL was also observed on decision accuracy $F(2, 81) = 4.13$, $p = .020$, $\eta_p^2 = .09$. Participants were significantly more accurate in low TL ($M = 4.98$, $SD = 1.02$) than moderate TL ($M = 4.12$, $SD = .89$), $p = .016$. No significant difference between high TL ($M = 4.48$, $SD = .44$) and low TL was observed, $p = .296$ or between high and moderate $p = .714$.

These findings suggest that the moderate TL impaired individuals' ability to make accurate decisions compared to the high and low TL conditions. Overall, the results demonstrate that participants made more accurate decisions in conditions of low TL.

In contradiction with HP4-A, no main effect of group was observed, F(2, 81) = .27, p = .767, η_p^2 = .01. In this task, playing video games did not impact on individual's decision accuracy. Further to this, no interaction effects were observed between TL and group F(4,162) = 1.35, p = .276, η_p^2 = .06. DC and TL F(4, 162) = 1.26, p = .289, η_p^2 = .03, DC and group F(4, 162) = .26, p = .900, η_p^2 = .01, nor DC, group and TL, F(8,162) = .67, p = .716, η_p^2 = .03.

In sum, the results show that decision accuracy was impaired in the decision events that were low in criticality. In addition, decision accuracy was also reduced in the moderate TL. Hence, both DC and TL impacted on decision accuracy. In this experiment, playing video games did not impact on decision accuracy.

Table 13: Means and Standard Deviations for Accuracy as influenced by DC, TL and group

TL	Group	Overall	High DC	Medium DC	Low DC
High	NG	13.70 (2.63)	4.90 (2.55)	4.60 (2.11)	4.20 (1.55)
	VGP1	14.30 (5.27)	5.30 (1.95)	5.00 (1.63)	4.00 (2.71)
	VGP2	12.40 (3.10)	5.10 (1.66)	3.40 (1.17)	3.80 (1.40)
	Total	13.47 (3.79)	5.10 (2.02)	4.33 (1.77)	4.00 (1.91)
Moderate	NG	12.40 (2.76)	4.40 (1.78	4.60 (2.06)	3.60 (2.01)
	VGP1	11.90 (3.90)	4.60 (1.58)	4.20 (1.87)	3.10 (1.60)
	VGP2	12.70 (2.16)	5.10 (1.10)	4.90 (1.10)	2.60 (2.12)

	Total	12.33 (3.94)	4.70 (1.49)	4.57 (1.70)	3.10 (1.90)
Low	NG	15.00 (4.16)	6.20 (2.04)	5.10 (2.13)	3.70 (1.24)
	VGP1	13.40 (3.03)	5.40 (1.42)	4.40 (1.37)	3.60 (1.90)
	VGP2	16.10 (3.70)	6.40 (2.32)	5.40 (1.65)	4.60 (2.12)
	Total	14.83 (3.71)	6.00 (1.95)	4.97 (1.71)	3.97 (1.79)
Total	NG	13.70 (3.32)	5.17 (2.21)	4.77 (2.05)	3.83 (1.60)
	VGP1	12.30 (4.15)	5.10 (1.65)	4.53 (1.59)	3.57 (2.08)
	VGP2	13.73 (3.40)	5.53 (1.81)	4.57 (1.55)	3.67 (2.02)
Total		13.54 (3.61)	5.26 (1.89)	4.62 (1.73)	3.69 (1.89)

Note. Standard deviations are in parenthesis.

5.4.2. Confidence:

Participants were asked to rate confidence in their decision ‘0’ being *not confident at all* and ‘5’ *being extremely confident*. The maximum confidence score in total was 150 and for each DC 50. A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data. Mauchly’s test of sphericity was found to be significant as such Greenhouse-Geisser is reported (see Table 14).

A main effect of DC was found on decision confidence F(1.7, 135.3) = 7.56, p = .002, η_p^2 = .09. Bonferroni corrected post hoc tests showed that participants were significantly more confident in low DC events (M = 37.82, *SD* = 8.85) than medium DC events (*M* = 36.14, *SD* = 7.42), *p* = .022 and also more confident in high DC (*M* = 38.73, *SD* = 8.88) than medium DC, *p* < .001. No differences were found between high DC and low DC, *p* = .794. These finding demonstrate that participants were equally confident in high and low DC events and found medium DC events reduced confidence.

Levene's test significant was found to be significant. Contrary to HP2-C, confidence was not influenced by TL $F(2, 81) = 1.75$, $p = .180$, $\eta_p^2 = .04$. Hence, confidence remained the same across all TL conditions.

There was however a main effect on group in reported confidence $F(2, 81) = 8.10$, $p = .001$, $\eta_p^2 = .17$. Post hoc comparisons revealed that VGP2 were significantly more confident (*M* = 40.23, *SD* = 6.27) in their decisions than NG (*M* = 35.33, *SD* = 12.38), *p* < .001. No significant differences were found between NG and VGP1, *p* = .242 or VGP1 and VGP2, *p* = .083. As such, in support of HP4-B, individuals who play more hours video games place more confidence in their decisions.

Further, no interactions were observed close to significant interaction between DC and TL $F(3.34, 135.30) = 2.45$, $p = .060$, $\eta_p^2 = .06$, no significant findings between DC and group, $F(3.34,135.30) = 2.09$, $p = .097$, $\eta_p^2 = .05$, DC, group, TL $F(6.68, 135.30) = .71$, $p = .656$, $\eta_p^2 = .03$ or group and TL $F(4,81) = .05$, $p = .996$, $\eta_p^2 = .00$.

In sum, decision confidence was influence by DC and group. Low DC produced higher decision confidence but, both high and low DC decisions were made with higher confidence in comparison to those of medium DC. Additionally, VGP2s were more confident in their decisions in this task.

Table 14: Means and Standard Deviations for Confidence as influenced by DC, TL and group

TL	Group	Overall	High DC	Medium DC	Low DC
High	NG	93.80 (32.24)	31.10 (11.48)	30.00 (10.78)	36.20 (18.78)
	VGP1	107.90 (18.51)	33.00 (12.59)	34.70 (5.93)	36.70 (6.17)

	VGP2	117.50 (14.71)	41.70 (12.95)	38.90 (6.71)	38.40 (5.48)
	Total	106.40 (24.37)	35.27 (11.10)	34.53 (8.64)	37.10 (11.47)
Moderate	NG	102.30 (24.10)	35.60 (7.56)	32.90 (7.62)	34.30 (10.85)
	VGP1	115.20 (19.43)	40.20 (7.25)	37.20 (6.11)	37.80 (6.53)
	VGP2	126.40 (10.39)	44.50 (3.47)	41.20 (2.74)	40.20 (7.12)
	Total	114.63 (20.74)	40.10 (7.17)	37.10 (6.62)	37.43 (8.46)
Low	NG	106.40 (12.64)	37.00 (4.97)	33.90 (4.63)	35.50 (4.60)
	VGP1	116.10 (17.68)	40.30 (1.13)	36.40 (6.11)	39.20 (5.43)
	VGP2	119.10 (20.31)	45.20 (6.55)	40.10 (8.36)	42.10 (6.19)
	Total	113.87 (17.47)	40.83 (7.96)	36.80 (6.83)	38.93 (5.93)
Total	NG	100.83 (24.10)	34.57 (8.54)	36.27 (7.97)	34.57 (8.54)
	VGP1	113.07 (18.26)	37.83 (9.66)	36.10 (5.39)	37.90 (5.94)
	VGP2	121.00 (15.63)	43.80 (5.51)	40.07 (6.24)	40.23 (6.27)
Total		111.63 (21.14)	38.73 (8.88)	36.14 (7.42)	37.82 (8.85)

Note. Standard deviations are in parenthesis.

5.4.3. W-S C-A:

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data between TL, DC and group on individuals within-subjects confidence-accuracy (W-S C-A). No significant main effect of DC shown, F(2, 162) = .53, p = .590, η_p^2 = .01. There was also no main effect of TL on W-S C-A F(2, 81) = 1.39, p = .254, η_p^2 = .03 and no main effect of group F(2, 81) = .32, p = .730, η_p^2 = .01. Further, no interaction was observed between DC and TL F(4, 162) = .84, p =.505, η_p^2 = .02. No interaction was observed between DC and group F(4, 162) = .47, p = .761, η_p^2 = .01. No interaction between DC, TL and group F(8, 162) = .50, p = .857, η_p^2 = .02. No interaction between TL and group F(4, 81) = .58, p = .681, η_p^2 = .03. This finding demonstrates that W-S C-A was not influenced by TL, DC or group. In sum, overall W-S C-A scores were very low however, not negative (M = .03, SD = .35). Due to the lack of significant findings, the results table reporting the means and standard deviations of W-S C-A as influenced by DC and TL are reported in Appendix 10a.

5.4.4. Percentage Confidence in Correct and Incorrect responses:

W-S C-A demonstrates the relationship between confidence and accuracy. However, a high correlation suggests both being highly confident in correct decisions as well as low confidence in incorrect decisions. Similary, a negative correlation would suggest that individuals are highly confident in incorrect reposnses or not confident in correct responses. By examining the percenatge confidence in incorrect or correct responses, the direction of the confidence (over/under confidence) can be displayed.

To look at some of the variations in the data and examine the zero correlation found in W-S C-A, percentage confidence in correct and incorrect responses was calculated. To do this the number of correct responses was recorded and the confidence in those decisions calculated to produce a confidence percentage in correct responses. The same was done for incorrect responses (Table 15). Interestingly all data suggests a high degree of confidence in decisions

To examine whether there were any differences in percenatage confidence an 3 X 3 ANOVA was firstly conducted on group, condition and percentage confidence in correct responses.

Levene's was found to be significant. A main effect of TL was found F(2,81) = 3.37, p = .039, η_p^2 = .08. Individuals were significantly more confident in their correct responses in the low TL (*M* = 78.90, *SD* = 12.37) than they were in the high TL (*M* = 70.21, *SD* = 15.61), p = .036. No differences between high and moderate (*M* = 75.70, *SD* = 13.55), p = .326 and no differences between low and moderate TL, p = 1.00.

A main effect of group was also observed F(2,81) = 8.31, p = .001, η_p^2 = .17. VGP2 were significantly more confident in their correct decisions (*M* = 81.85, *SD* = 11.08) than NG (*M* = 68.07, *SD* = 15.86), $p < .001$. No differences between NGs and VGP1(*M* = 74.90, SD = 12.16), p = .140. or VGP1 and VGP2, p = .129. No interaction effect was observed F(4.81) = .15, p = .965, η_p^2 = .01.

Levene's test was found to be significant. Similary, a 3 X 3 ANOVA was conducted on group and condition and percentage confidence in incorrect responses. No main effect of TL was observed F(2,81) = 1.72, p = .185, η_p^2 = .04. A main effect of group was found F(2,81) = 7.61, p = .001, η_p^2 = .16. Again, VGP2 were significantly more confident in incorrect responses (*M* = 81.01, *SD* = 11.45) than NG (*M* = 67.03, *SD* = 16.54), p = .001. No differences between NGs and VGP1 (*M* = 74.93, *SD* = 12.70), p = .092, or VGP1 and VGP2, p = .284. No interaction effect was observed F(4,81) = .09, p = .987, η_p^2 = .00.

In general, individuals displayed the same percentage of confidence in correct and incorrect responses (*M* = 74.94, *SD* = 14.22) and (*M* = 74.33, *SD* = 14.75) respectively. Thus, suggesting that individuals cannot differentiate between correct and incorrect responses.

Table 15: Means and Standard Deviations for % confidence (correct and incorrect) according to TL and group

TL	Group	% Correct	% Incorrect
High	NG	62.90 (21.24)	61.80 (21.73)
	VGP1	70.30 (10.31)	71.60 (14.01)
	VGP2	77.45 (10.51)	78.03 (11.64)
	Total	70.22 (15.61)	70.48 (17.19)
Mod	NG	67.30 (15.19)	68.70 (17.19)
	VGP1	76.20 (12.76)	76.90 (13.54)
	VGP2	83.60 (7.09)	82.90 (8.35)
	Total	75.70 (13.55)	76.17 (14.33)
Low	NG	74.00 (7.69)	70.60 (8.14)
	VGP1	78.20 (13.04)	76.30 (11.02)
	VGP2	84.50 (14.21)	82.10 (14.23)
	Total	78.90 (12.37)	76.33 (12.00)
Total	NG	68.07 (15.86)	67.03 (16.54)
	VGP1	74.90 (12.16)	74.93 (12.70)

VGP2	81.85 (11.08)	81.01 (11.46)
Total	74.94 (14.22)	74.33 (14.75)

Note. Standard Deviations are in parenthesis

5.4.5. WL and SA:

Manipulation check: A 3 X 3 ANOVA was conducted on TL, group and reported overall SA. *Levene's test was found to be significant.* However, a significant main effect of TL was observed $F(2, 81) = 6.32$, $p = .003$, $\eta_p^2 = .16$. Post hoc test revealed that individuals reported higher levels of SA in low TL (*M* = 20.1 *SD* = 1.03) than high TL (*M* = 14.8, *SD* = 1.03) $p = .001$. No main effect of group $F(2,81) = .641$, $p = .529$, $\eta_p^2 = .02$ and no interaction effect $F(4,81) = .38$, $p = .822$, $\eta_p^2 = .02$ were observed. These findings suggest that the low TL condition allowed individuals to feel that had a better understanding of the situation as reported by higher SA scores.

Manipulation check: A 3 X 3 ANOVA was also carried out on TL, group and WL. A main effect of TL was found $F(2, 81) = 10.23$, $p < .001$, $\eta_p^2 = .20$. Significant differences between high and low TL were found. Those in the high TL reported significantly higher WL (*M* = 65.22, *SD* = 2.67) compared to the low TL (*M* = 48.33, *SD* = 2.66), $p < .001$. These findings suggest the TL manipulation was successful. No main effect of group $F(2, 81) = 1.56$, $p = .216$, $\eta_p^2 = .04$ and no interactions were found $F(4, 81) =.57$, $p = .686$, $\eta_p^2 = .03$. In sum, high TL increased subjective feelings of WL during the task.

Table 16: Means and Standard Deviations for WL and SA as influenced by TL and group

TL	Group	Overall SA	Overall WL
High	NG	13.70 (6.06)	65.28 (14.05)
	VGP1	13.60 (4.22)	62.90 (16.23)
	VGP2	17.10 (5.44)	67.48 (7.16)
	Total	14.80 (5.37)	65.22 (12.75)
Mod	NG	16.90 (8.89)	63.50 (16.64)
	VGP1	18.20 (8.22)	56.94 (14.82)
	VGP2	17.50 (4.60)	55.81 (15.03)
	Total	17.53 (7.23)	58.75 (15.37)
Low	NG	19.90 (3.81)	55.04 (13.95)
	VGP1	19.70 (4.57)	46.11 (11.33)
	VGP2	20.70 (3.56)	43.84 (18.96)
	Total	20.10 (3.89)	48.33 (15.36)

	NG	16.83 (6.86)	61.27 (15.11)
Total	VGP1	17.17 (6.32)	55.32 (15.48)
	VGP2	18.43 (4.73)	55.71 (17.15)
	Total	17.48 (6.01)	57.43 (15.99)

Note. Standard Deviations are in parenthesis

5.4.6. SA Subscales:

The SA measure of SART is made up of attentional supply, demand and understanding. One way ANOVAs were conducted on these subscales to examine the effect of group and TL (see Table 17).

5.4.6.1 Demand: A main effect of TL was found in the demand subscale. High and moderate TL conditions were found to be significantly more attentionally demanding than the low TL condition, $F(2, 81) = 8.50$, $p < .001$, $\eta_p^2 = .17$. Post hoc tests revealed significant differences between High TL (*M* = 13.90, *SD* = 2.81) and low TL (*M* = 10.17, *SD* = 3.62), $p = .001$, and between moderate TL (*M* = 13.37, *SD* = 4.50) and low TL, $p = .005$. Therefore, individuals found the task to be high on instability, variability and complexity in both the high and moderate TL conditions. No main effect of group $F(2, 81) = .22$, $p = .801$, $\eta_p^2 = .01$, and no interaction between group and TL was found, $F(4, 81) = .32$, $p = .866$, $\eta_p^2 = .02$.

5.4.6.2. Supply: Significant differences between attentional supply were also found. There was a main effect of TL $F(2,81) = 3.54$, $p = .034$, $\eta_p^2 = .08$. Post hoc tests reveal a close to significance between moderate and low TL, $p = .059$. Individuals reported higher levels of supply in the moderate TL (*M* = 19.27, *SD* = 3.72) compared to the low TL (*M* = 17.23, *SD* = 2.93). No significant difference between high and moderate or high, $p = .087$ and low, *p* = 1.00. These findings suggest that the moderate TL condition produced higher feelings of arousal, lower spare mental capacity and more division of attention. No main effect of group F(2,81) = .02, p = .984, η_p^2 = .00 and no interaction effect between group and TL F(4, 81) = .63, p = .642, η_p^2 = .03 were observed.

5.4.6.3. Understanding: No main effects for understanding was found, F(2,81) = 1.83, p = .167, η_p^2= . 04. No main effect of group existed, F(2,81) = 2.01, p = .140, η_p^2 = .05 and no interaction was displayed between TL and group F(4, 81) = 1.22, p = .309, η_p^2 = .06. The findings suggest all conditions were rated at reasonably equivalent levels of understanding.

Table 17: Means and Standard Deviations for dimensions of SA as influenced by TL and group

TL	Group	Demand	Supply	Understanding
High	NG	13.50 (3.63)	16.50 (3.03)	10.70 (2.58)
	VGP1	14.10 (1.97)	17.30 (3.23)	10.40 (3.06)
	VGP2	14.10 (2.85)	18.30 (2.83)	12.90 (4.12)
	Total	13.90 (2.81)	17.37 (3.02)	11.33 (3.40)
Moderate	NG	14.30 (5.00)	19.10 (3.78)	12.10 (3.70)
	VGP1	12.70 (4.24)	19.60 (3.57)	12.50 (2.76)
	VGP2	13.10 (4.55)	19.10 (4.18)	11.50 (4.88)
	Total	13.37 (4.50)	19.27 (3.72)	12.03 (3.76)
Low	NG	10.20 (4.05)	18.00 (2.45)	11.90 (3.28)
	VGP1	9.50 (4.09)	17.10 (2.64)	12.10 (2.81)

	VGP2	10.80 (2.86)	16.60 (3.69)	15.10 (3.12)
	Total	10.17 (3.62)	17.23 (2.93)	13.03 (3.38)
Total	NG	12.67 (4.90)	17.87 (3.21)	11.57 (3.17)
	VGP1	12.10 (3.98)	18.00 (3.27)	11.67 (2.93)
	VGP2	12.67 (3.67)	18.00 (3.64)	13.17 (4.28)
	Total	12.48 (4.02)	17.96 (3.34)	12.13 (3.55)

Note. Standard Deviations are in parenthesis

5.4.7. NASA TLX Subscales:

There are 6 Subscales of WL as measured by NASA TLX (mental demand, physical demand, temporal demand, performance, effort and frustration). To identify a specific type of workload ANOVAs were conducted to examine these subscales by TL and group (see Table 18).

5.4.7.1. Temporal Demand: Significant differences were found between TL conditions in temporal demand, F(2, 81) = 8.53, $p < .001$, η_p^2 = .17. Hence, participants found the high TL condition (*M* = 278.17, *SD* = 108.17) more temporally demanding than the low TL condition (*M* = 160.67, *SD* = 103.70), $p < .001$. No differences between high and moderate, *p* = .066 or moderate and low, *p* = .235 were found. Furthermore, no main effect of group, F(2, 81) = 1.16, *p* = .320, η_p^2 = .03 or interaction was observed, F(4,81) = 1.42, *p* = .234, η_p^2 = .07.

5.4.7.2. Effort: A significant main effect of TL was also found for effort, F(2, 81) = 3.78, *p* = .027, η_p^2 = .09. Participants reported putting more energy and effort into the high TL condition (*M* = 161.50, *SD* = 105.84) than the low TL condition (*M* = 101.33, *SD* = 79.03), *p* = .033. No differences, however were found between high and moderate TL (*p* = 1.00) and moderate and low TL, *p* = .124. No main effect of group was observed, F(2, 81) = 2.18, *p* = .119, η_p^2 = .05 and no interaction was shown, F(4, 81) =.77, *p* = .549, η_p^2 = .04. Moreover, no other dimensions of WL were found to be significant.

Table 18: Means and Standard Deviations for dimensions of WL as influenced by TL and group

TL	Group	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
High	NG	261.50 (122.39)	0.50 (1.58)	335.00 (116.17)	154.00 (121.01)	145.50 (91.18)	85.00 (96.26)
	VGP1	276.50 (126.69)	1.00 (3.16)	267.00 (58.18)	152.50 (112.40)	135.00 (91.86)	112.00 (82.77)
	VGP2	306.00 (80.86)	8.00 (18.29)	232.50 (121.57)	117.50 (58.94)	204.00 (127.56)	144.50 (138.77)
	Total	281.33 (109.60)	3.17 (10.95)	278.17 (108.17)	141.33 (99.19)	161.50 (105.84)	113.83 (107.66)
Mod	NG	276.50 (115.01)	13.00 (34.34)	180.00 (124.92)	178.00 (126.06)	183.00 (108.35)	135.00 (140.83)
	VGP1	221.00 (103.25)	0.00 (0.00)	245.00 (122.11)	132.50 (85.58)	122.00 (62.55)	134.00 (126.36)
	VGP2	194.00 (114.18)	9.00 (28.46)	209.50 (124.31)	121.50 (114.02)	143.00 (70.29)	160.50 (154.57)
	Total	230.50 (112.61)	7.33 (25.45)	211.50 (122.46)	144.00 (108.89)	149.33 (83.98)	143.17 (136.68)
Low	NG	223.00 (127.50)	0.00 (0.00)	199.00 (120.71)	175.50 (93.44)	116.50 (93.45)	115.00 (121.36)
	VGP1	230.50 (150.95)	6.50 (18.86)	141.00 (80.03)	156.00 (80.06)	72.00 (57.70)	88.00 (63.79)
	VGP2	190.00 (115.71)	1.50 (4.74)	142.00 (105.94)	114.10 (113.17)	115.50 (81.39)	95.00 (107.13)
	Total	214.50 (128.81)	2.67 (11.20)	160.67 (103.70)	148.53 (96.71)	101.33 (79.02)	99.33 (97.63)
Total	NG	253.67 (119.68)	4.50 (20.10)	238.00 (135.95)	169.17 (110.93)	148.33 (98.49)	111.67 (118.48)

VGP1	242.67 (126.37)	2.50 (11.05)	217.67 (103.87)	147.00 (91.07)	109.67 (75.03)	111.33 (93.32)
VGP2	230.00 (114.98)	6.17 (19.33)	194.67 (119.96)	117.70 (95.38)	154.17 (100.27)	133.33 (133.26)
Total	242.11 (119.47)	4.39 (17.19)	216.78 (120.60)	144.62 (100.63)	137.39 (93.09)	118.78 (115.35)

Note. Standard deviations are in parenthesis.

5.4.8 Between-Subjects Confidence- Accuracy:

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was conducted. It was found that overall accuracy was not related to overall confidence $r(88) = -.13$, $p = .242$.

5.4.9. Relationship between WL, SA, Accuracy, Confidence and W-S C-A:

To assess the relationship between WL and SA a series of Pearson's correlations were calculated. A correction for multiple comparisons was applied, the level to reach significance is $p = .005$. Overall confidence in decisions was significantly and moderately related to overall feelings of WL, $r(88) = -.32$, $p = .002$. Therefore, WL had a negative impact on confidence in decisions, with higher levels of WL reducing an individual's confidence in decisions.

Overall confidence was also significantly and moderately related to overall SA, $r(88) = .49$, $p < .001$. As confidence increased, reported levels of SA were higher. Individuals who felt that they had more of an awareness of the situation were more confident overall in their decisions. No significant relationships were found between SA and accuracy or WL and accuracy in decisions, $p > .005$. No significant relationships between W-S C-A and SA and WL. In sum, confidence was influenced by both SA and WL.

5.4.10. Personality Constructs:

To assess and compare each group (i.e. NG, VGP1, VGP2) on the psychometric measures (Extraversion, agreeableness, openness to experience, conscientiousness, neuroticism), one way ANOVAs were performed on the data. A significant difference was found between the groups in the trait of conscientiousness, $F(2, 89) = 9.98$, $p < .001$. VGP2 players were significantly less conscientious ($M = 24.53$, $SD = 8.44$) than both NG ($M = 33.47$, $SD = 6.47$), $p < .001$ and VGP1 ($M = 30.77$, $SD = 8.73$), $p = .009$ No other comparisons for personality measures were significant, $p > .05$.

Again, Pearson's correlations were conducted to establish whether accuracy, confidence, and W-S C-A were related to cognitive constructs. Accuracy was significantly negatively related to ambiguity $r(88) = -.34$, $p < .001$ and decision style $r(88) = -.39$, $p < .001$. With the multiple comparison correction applied a close to significant finding was observed in the measure of decisiveness $r(88) = -.25$, $p = .019$. Individuals who scored more highly on these measures achieved lower accuracy scores. No other relationships were found to be significant, $p > .005$.

Table 19: Means and Standard Deviations for dimensions of psychometric scores as influenced by group

Measure	NG	VGP1	VGP2	Total
Ambiguity Tolerance A	49.60 (6.34)	49.17 (8.95)	50.63 (7.36)	49.80 (7.53)
Decisiveness	23.07 (4.59)	20.53 (5.51)	21.70 (5.83)	21.77 (5.45)
Ambiguity Tolerance B	37.23 (7.56)	35.05 (6.69)	36.80 (6.34)	36.37 (6.87)
Decision Style	61.60 (10.93)	55.40 (10.65)	57.90 (11.09)	58.30 (11.07)
Neuroticism	24.50 (9.72)	21.40 (10.07)	23.13 (8.88)	23.01 (9.54)
Extraversion	30.03 (7.12)	29.43 (7.74)	26.80 (7.99)	28.76 (7.67)

Openness To experience	31.90 (5.69)	33.47 (6.39)	32.47 (5.49)	32.61 (5.84)
Agreeableness	30.80 (6.03)	31.70 (7.31)	29.90 (7.04)	30.80 (6.78)
Conscientiousness	33.47 (6.47)	30.77 (8.73)	24.53 (8.44)	29.59 (8.71)

Note. Standard deviations are in parenthesis.

5.5. Discussion

The aim of this first experiment was to begin to establish the external (DC, TL) and internal factors (group, personality and cognitive constructs) on measures of accuracy, confidence, and W-S C-A. WL and SA were also assessed in relation to performance measures and task manipulation checks. Overall, the experiment found that DC impacted on both decision accuracy and confidence. Decision confidence increased in low DC events, in comparison, decision accuracy increased in high DC events. TL in this experiment was also found to influence decision accuracy. However, decision confidence was unrelated to TL. Metacognition, as measured by W-S C-A was low and unrelated to any of the variables. Cognitive constructs which related to dealing with uncertainty, were also found to be related to decision accuracy in this study. In addition, VGPs demonstrated higher levels of decision confidence. Differences in WL and SA were found between the TL conditions which suggest the TL manipulation was successful.

5.5.1. Accuracy:

The analysis showed DC impacted on the accuracy of decisions. Individuals made more accurate decisions in high DC events. Thus, making fewer errors in decisions that held higher consequences, supporting the findings from Adams-White et al. (2018). Moreover, this outcome also supports previous literature that criticality influences performance (Hanson et al., 2014; Wheatcroft et al., 2017). However, contrary to the HP1-A, individuals did not make more accurate decisions in lower levels of criticality (Bliss & McAbee, 1995). Such a result may be explained by the fact that research has shown that task performance increases when participants find the task more important perhaps indicating that individuals believed high DC decisions to be important within the task context. However, research has also found that criticality may have a positive impact on performance (Callister et al., 1999). This could also relate to participants applying more attention and effort to decisions that held a greater consequence for an incorrect decision. Furthermore, in this experiment participants were not informed about the different decision criticalities. As such, this finding relates to the individuals perception of decision criticality, in comparisons to the previous studies in which participants were informed on criticality (Bliss & McAbee, 1995; Hanson et al., 2014). Additionally, this experiment aimed to increase understanding of DC by examining individual decision events which varied in criticality (high, medium, low) as opposed to the criticality of a whole scenario. As previously demonstrated the decision event itself is important to consider (Wheatcroft et al., 2017; Adams-White et al., 2018). These findings demonstrate that the criticality of each decision influences the accuracy of decisions. Hence, future research should continue to examine the criticality of decision events - not just the criticality of a scenario.

In support of HP2-A, which stated that high TL would reduce decision accuracy, individuals were more accurate when the TL was low. In this study, as described in Chapter two, TL related to cognitive load and temporal demand imposed in each condition. As such, the low TL was characterised by low cognitive load on the operator and reduced temporal demand in the task. These findings provide support to other research which has found that cognitive load impairs decision accuracy (Cumming & Harris, 2001; Keinan et al., 1987). The finding is further supported by the analysis of the task manipulation of WL. The outcomes found that overall WL, and the subscales of temporal demand and effort, showed increased levels in the high TL compared to the low TL. This would suggest that individuals felt that they had to put more effort into the task when the TL was high. Effort referred to how hard the individual felt they had to work in order to achieve a level of performance. An implication of this finding is the possibility that, although individuals put more effort into the task, this did not translate to the individual being more accurate. Consequently, this would also suggest that individuals believed themselves to be performing better than they actually were.

With regards to differences in group accuracy scores, as mentioned in Chapter two, previous research has highlighted that VGPs may be better suited to making decisions as they may possess a range of cognitive skills that may benefit them in a decision making task (Spence & Feng, 2010; McKinley et al., 2011; Wheatcroft et al., 2017). This experiment, therefore, aimed to examine different populations, comparing gamers, which were made up of those who played more than 7 hours a week (VGP1) and those who played up to 7 hours a week (VGP2), with Non-Gamers (NG). As there is a dearth of research on gamers and decision accuracy an exploratory hypothesis (HP4) stated that there would be differences between the groups. However, the results from this study found no differences between groups in regards to their decision accuracy. As a result, in this task, playing video games did not impact on the accuracy of individuals decisions. This could be explained by the novelty of this task for both gamers and NGs which would also suggest that domain-specific experience is relevant to decision accuracy (Nicolson & O'Hare, 2014).

5.5.2. Confidence:

In keeping with HP1-B, DC was also found to impact on individual confidence. Participants were more confident in their decisions when they were presented with either a highly critical decision or a low critical decision, in comparison to low confidence displayed in the medium DC decisions. One explanation might be that the medium decisions created higher levels of uncertainty and confusion, thus causing a decrease in confidence levels. Indeed, uncertainty has been shown to decrease decision confidence (Heerman & Walla, 2011). Other research in different contexts has also shown that the difficulties of items can impact on decision confidence (Kebbell, Wagstaff & Covey, 1996).

However, the results from this study did not find any evidence to support HP2-B which stated that decision confidence was influenced by TL. In this experiment, decision confidence remained consistent across all TL conditions. This supports previous literature that confidence may be a stable individual tendency and therefore confidence is not necessarily influenced by the demands of the task (Pallier et al., 2002; Burns, Burns & Ward, 2016). This is further supported by the findings that participants remained confident regardless of their experience of WL. As mentioned, WL was significantly higher in the high TL condition and crucially, decision accuracy was also impaired in the high TL condition. Thus, suggesting that individuals may be unaware of the impact of WL on their decision making ability. It is important that operators know when they are overloaded to ensure effective decision making and the ability to apply the correct amount of confidence to these decisions. For instance, overload has been associated increasing the time to make a decision and uncertainty (Cohen, 1980). This finding suggests there is a need to improve understanding of overconfidence in critical environments and impact on decision making.

The results in this experiment also found that VGP2 were significantly more confident in their decisions than both NG and VGP1, thus supporting Wheatcroft et al. (2017) finding that VGPs demonstrated significantly higher levels of confidence. One explanation for this finding could be that experience, such as playing games, can result in increased decision confidence (Wheatcroft et al., 2017; Atinaja-Faller et al., 2010; Chung & Monroe, 2000). Furthermore, familiarity can also result in increased decision confidence by providing a belief that individuals are accurately remembering important detail (Wheatcroft et al., 2017; Chandler, 1994). However, and crucially, VGP2 confidence was not linked to the accuracy of their decisions. Therefore, unlike the outcomes found in Wheatcroft et al. (2017) gamers in this study were not more suited to make confident and accurate decisions. However, Wheatcroft et al. (2017) conducted this experiment in an air domain. As mentioned, it could, therefore, support the argument that task-specific experience is important to decision making, and not just general gaming experience. Furthermore, no significant differences were found for VGP1 which suggests the amount of time playing games influences confidence in decisions and not necessarily just playing video games. There is currently, a dearth of research on the impact of the amount of time spent playing games. One study found that gamers who play more hours of games have a tendency to make riskier decisions (Bailey, West & Kuffel, 2013). Hence, this overconfidence could be linked to the tendency for gamers to take more risks.

5.5.3. W-S C-A:

Moving on to the relationship between decision confidence and accuracy as measured by W-S C-A. As mentioned, a good relationship between confidence and accuracy is required for optimal decision making as it demonstrates an individual’s ability to apply the appropriate levels of confidence to corresponding correct or incorrect decisions (Wheatcroft & Woods, 2010). However, as yet, there is a lack of research into this relationship in the context of air defence decision making (Adams-White et al., 2018). As such this experiment aimed to further establish whether this relationship is influenced by any of the variables related to air defence decision making.

Contrary to expectations, but in agreement with Adams-White et al. (2018), no significant differences were found between W-S C-A. That is to say that DC, TL nor group had an impact of individual’s ability to discriminate between accurate and inaccurate responses. Furthermore, in this study, the overall correlations between confidence and accuracy were low and close the zero. The absence of a relationship between confidence and accuracy could be explained by the lack of experience of the population in this study. Research suggests that training and experience can increase calibration (Lichtenstein, Fischhoff & Phillips, 1977). There has been research to suggest that there is a general confidence factor (Kleitman &

Stankov, 2007) and individuals have a habitual way in which they assess the accuracy of their decisions (Stankov, Kleitman & Jackson, 2015). This is supported by the findings in this study that no significant differences were found between TL conditions. In addition, no relationship was found between the accuracy of decisions and confidence (Between-Subjects Confidence-Accuracy) expressed in these decisions was found. This finding supports previous research which has repeatedly demonstrated that confidence is not a good indicator of accuracy (Deffenbacher, 1980).

5.5.4. WL and SA:

As stated in HP-5, this study was also interested in increasing understanding of decision making against already established performance measures such as SA and WL. These measures also provided an indication of whether the manipulation of TL was successful. SA has been argued to be important to military decision making however, there is little research on how this directly related to decision making ability and the potential confounding variables measured in this experiment (DC, TL, individual differences). In this study higher SA was reported in low TL. As such, these findings suggest that the low TL condition allowed individuals to feel that had a better understanding of the situation. Hence, similarly to accuracy, individuals may have had more cognitive resources available to assess the situation. However, SA is not equivalent to performance (Stanners & French, 2005) further supported by the finding in this study that SA was not directly related to decision accuracy.

Additionally, no differences in global SA were found between the groups. This finding suggests that although VGP2s were more confident in their decisions, they were not more confident in their assessment of SA. Supporting the findings of Vidulich et al. (1995) who found that video game experience did not improve SA. Another explanation could be that SA is domain specific and although video game experience is beneficial to some aspects of cognitive ability and decision making, again suggesting that the type of experience is important (Nicolson & O’Hare, 2014).

Furthermore, analysis into differences in the subscales of SA found that high and moderate TL conditions were found to be significantly higher on the attentional demand scale than the low TL condition. Therefore, individuals found the task in these two conditions to be high on instability, variability and complexity. Furthermore, individuals also reported higher levels of supply in the moderate TL compared to the low TL. These findings suggest that the moderate TL condition produced higher feelings of arousal, lower spare mental capacity and greater division of attention. On the other hand, all TL conditions were rated at reasonably equivalent on levels of understanding which would seem plausible as all participants were provided with the same information. These findings, therefore, suggest that demand and supply are influenced by the task manipulation of TL.

Unlike the findings of Stanners and French (2005), SA was not related to decision accuracy. However, in support of the findings of Adams-White et al. (2018), SA was found to be related to decision confidence. Individuals who reported higher levels of SA were also more confident in their decisions. Nevertheless, these findings should be taken with caution. SA was measured subjectively and a confidence bias has previously been found in SA reporting (Sulistyawati & Chui, 2009). Thus, it might be that individuals are generally confident in their assessments of SA performance. Importantly, SA was not related to accuracy in decisions taken. Individuals may privately believe they had a better understanding of the situation than they actually accepted. This is particularly salient as SA was not related to decision accuracy. In support, Endsley (1995) argued that self-ratings of SA tend to be related to a statement of how certain the person feels about SA.

In regards to WL, individuals in the high TL reported significantly higher feelings of WL compared to the low TL. These findings suggest the TL manipulation was successful as the increased frequency of decisions and frequency of decisions induced feelings of higher WL. The outcomes supporting the findings of Adams-White et al. (2018) and Loft and Sadler (2015) that WL increases with TL. In respect of increases in cognitive load in particular, and by analysing the subsections of WL, it was found that participants experienced the high TL condition more temporally demanding than the low TL condition. This suggests that the WL, in this instance, was most influenced by the speed at which the task occurred. Furthermore, participants reported putting more effort into the high TL condition than low TL condition. In this experiment, video game experience did not impact reported WL.

WL was found to be negatively related to overall decision confidence. Higher levels of WL resulted in lower levels of decision confidence. This is an important finding for decision making as reduced confidence in decisions taken could lead to increased WL. Individuals may seek out more information to support and/or negate decision certainty. No differences were found for accuracy and W-S C-A. Contrary to the findings of Kim et al. (2018) who found a negative relationship between metacognition and WL, in this experiment, W-S C-A was not related to WL. In sum, the findings of WL and SA may be useful for the design of experimental studies. Indeed, WL and SA have been found previously to be sensitive to variations in TL and that changes in task difficulty are a function of both WL and SA (Selcon et al., 1991).

5.5.5. Personality and Cognitive Constructs:

The first experiment also examined the individual differences in decision makers. Personality and cognitive constructs were used to assess whether any of the factors measured related to decision accuracy, confidence, W-S C-A, as well as, WL and SA. The investigations into broad personality constructs (Extraversion, Neuroticism, Openness to Experience, Agreeableness and Conscientiousness) in this experiment were found to be unrelated to confidence, accuracy, W-S C-A, or SA. This suggests the constructs may not be related to these measures or sufficiently salient to decision making processes. Thus, supporting previous research that personality is not related to performance (Wallenius et al., 2014).

However, unlike personality, measures of cognitive constructs were found to impact on decision accuracy. This experiment found that decision accuracy was negatively related to decisiveness, ambiguity and decision style. As discussed in Chapter two, these measures are related to uncertainty. The Ops room is largely uncertain and operators must be able to deal with the uncertainty without diminishing performance. This study demonstrated that being able to deal with uncertainty is beneficial to performance. One explanation is that individuals who are less tolerant to ambiguity are less accurate as they see the task as threatening and thus more likely to give up (Budner, 1962). It is probable that a lack of tolerance hindered individual's ability to make accurate decisions. Further, although not replicated in this study, Wheatcroft et al. (2017) found an intolerance of ambiguity was negatively related to W-S C-A (i.e., a greater tolerance of ambiguous conditions was related to increased W-S C-A). As such it may seem likely that a high Tolerance to Ambiguity, decisiveness and decision style are relevant to accurate decision making in air defence.

The group analysis found that VGP2 were the least conscientious. Conscientiousness is characteristic of individuals who are more efficient and organised and desire to do well in the task (Costa & McCrae, 1992). Interestingly, however, VGP2 displayed overconfidence in this task. This desire to do well may have misled them to believe they were doing better than they were.

The research conducted in this section is not without its limitations. Although the study was initially interested in the air defence role and novice capacity to the task, it remains that one limitation of this study was the use of novice participants rather than experts. However, Chapter two discusses increasing internal validity with larger sample sizes. As suggested by (Hoffman & Klein, 2017) it may be beneficial to the NDM paradigm to gain an understanding of how expertise is developed. For instance, Klein, Hintze and Saab (2013) developed the ShadowBox technique which helps novices to understand the decision making processes of experts. Therefore, the use of novices in this experiment can also be considered more positively as it allows for a baseline comparison to help understand decision making in the kinds of critical environments explored in this paper. Further, the use of novices may be beneficial to understand the training needs of less experienced decision makers in order they may be considered as a potential resource. Indeed, Chapter ten has addressed the limitation of using novice participants by recruiting expert participants.

Moreover, in the standardisation of the study, participants were allocated 20 seconds to make a decision. It is possible the timeframe could have affected the processes that individuals used to make their decisions; though due to the nature of air defence decision making, it is realistic to expect operators to be under some time pressure during the circumstances that surround these types of decisions. Time pressure is examined further in Chapters six and seven. In addition, a further limitation is the practice effect potential of the participant trial which was conducted to minimise participant fatigue during the actual experiment. Indeed, some research has found that practice can degrade certain aspects of metacognitive performance (Jackson et al., 2014) and it was deemed important to consider this outcome in relation to the design of the trial element. To maintain consistency throughout the work some of the limitations will also be applicable to future studies and as such, will not be discussed again until Chapter thirteen. As such, this Chapter will draw together and address the limitations of the Experimental Work in greater detail.

5.6. Summary

In sum, the first experiment demonstrated that DC impacts on both decision confidence and accuracy. DC impacted on both decision accuracy and confidence. Decision confidence increased when individuals made Low DC decisions and decision accuracy increased when individuals made high DC decisions. TL in this experiment was also found to influence decision accuracy. However, decision confidence was unrelated to TL. These findings suggest that research should include the decision events and not just the criticality of a full scenario. These findings have implications of informing operators to be more aware that different levels of decision criticality may impact on their decision making. Ensure that in low criticality individuals may need to gather more information before deciding on a decision for instance and apply more appropriate levels of confidence to the lower criticality decisions. To increase the external validity of the task and examine the impact of time pressure on the task the next experiment reduced the time given to individuals to make a decision. This first study began to answer some of the research questions in this thesis. It demonstrated that decision confidence and accuracy can be influenced by DC and TL and individuals differences. Individual differences in dealing with uncertainty were found to impact on the accuracy of decisions. However, W-S C-A did not provide any significant results.

CHAPTER SIX

6. Investigating the Impact of Time Pressure on Decision Accuracy, Confidence and W-S C-A

6.1. Introduction

Many of the decisions that are made in an Ops room will be made under time pressure. Previous authors have argued that there is a dearth of literature which references a specific time that causes feelings of stress (Maule & Hockey, 2000). Hence, it is difficult for researchers to determine how to vary the time given. Consequently, it has been argued that time pressure can be operationalised by adopting a given time, which is any fraction of the usual time to complete the task (Corso & Lobler, 2011). In the first experiment, participants were given 20 seconds to make a decision and were able to make a decision in this time, which suggests that they were not time pressured. To assess how time pressure might influence these factors, the second experiment reduced the time to make a decision from 20 seconds to 10 seconds. In doing so, it was envisioned that this would increase the time pressure on participants. As such, understanding what impact time pressure has on decision accuracy, confidence and W-S C-A, particularly as research has not yet demonstrated a clear impact of time pressure on these factors. Recent research found no difference in time pressure, however time pressure has been found to interact with task difficulty (Yang et al., 2012). Gonzalez et al. (2005) argued there is little research in dynamic task environments, such as air defence, that has investigated the impact of time pressure. In light of this, research is needed in more dynamic and realistic scenarios to assess the relationship of decision accuracy and confidence. Furthermore, this reduction in time may also increase the feelings stress on the decision maker (Keinan et al., 1987) and as such increase feelings of WL during the task.

6.2. Hypotheses

- **HP6:** High DC will increase decision accuracy (A), reduce confidence (B) and W-S C-A (C).
- **HP7**: High TL will reduce decision accuracy (A), confidence (B) and W-S C-A (C).
- **HP8**: Individual differences will be found in decision accuracy (A), confidence (B) and W-S C-A (C).
- **HP9:** VGPs will be more accurate (A), confident (B) and display higher W-S C-A (C).
- **HP10:** To explore the relationship between accuracy, confidence, W-S C-A and measurements of WL and SA.

6.3. Participants

Ninety participants were recruited through opportunity sampling from the University of Liverpool. The participants consisted of 40 females and 50 males with a mean age of 23 years (*SD* = 6.56). The groups were divided into 3 groups using the same criteria described in section 4.3.3.2. Ethical approval and statistical criteria remained the same for this experiment (see Section 4.3.2 and 5.3).

6.4. Results

To assess the differences in means a number of statistical analyses were performed on the data for accuracy, confidence and W-S C-A using ANOVA. A manipulation check was carried out to assess the differences in TL (see analysis of Workload and Situational Awareness). No significant differences were found between the TL manipulation in WL, $p = .164$ or SA, $p = .077$ as such the task manipulation was not successful. An alpha level of .05 was used for all statistical tests unless stated otherwise.

6.4.1. Accuracy:

The accuracy of the decisions was decided on by the SMEs. When designing the decision log and generating the decision options one of the decision options was voted to be the best decision given the current situation. Participants scored '1' for a*n accurate response* or '0' for *an* *incorrect response*. The maximum total was 30 and the maximum mean for each DC was 10. To examine the mean differences between DC, TL and group in accuracy an ANOVA was conducted.

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data. As Mauchly's test of sphericity was found to be significant, the Greenhouse-Geisser estimate for *df* was used (see Table 20).

A main effect of DC was found on accuracy, F(2,148.42) = 39.59, $p < .001$, $\eta_p^2 = .33$. Bonferroni corrected post hoc tests showed participants were more accurate in high DC decisions (*M* = 5.52, *SD* = 1.56) than both medium DC events (*M* = 4.97, *SD* = 1.88), $p < .001$ and low DC (*M* = 3.57, *SD* = 1.73), $p = .016$. Additionally, participants were significantly more accurate in medium DC decisions than low DC decisions, $p < .001$. These findings show that low DC events impaired decision accuracy.

However, no main effect of TL was found on accuracy F(2, 81) = 1.16, $p = .32$, $\eta_p^2 = .03$. No main effect of group was found either F(2, 81) = .65, $p = .938$, $\eta_p^2 = .00$. However, a significant interaction effect was observed between group and TL F(4,81) = 3.31, $p = .015$, $\eta_p^2 = .14$. The interaction between group and TL showed that VGP1 were more accurate in the high TL compared to VGP2 who were more accurate in the low TL. NG were similar across TL conditions (see Figure 9). This could be explained that having less familiarity to a computer-based gaming tasks.

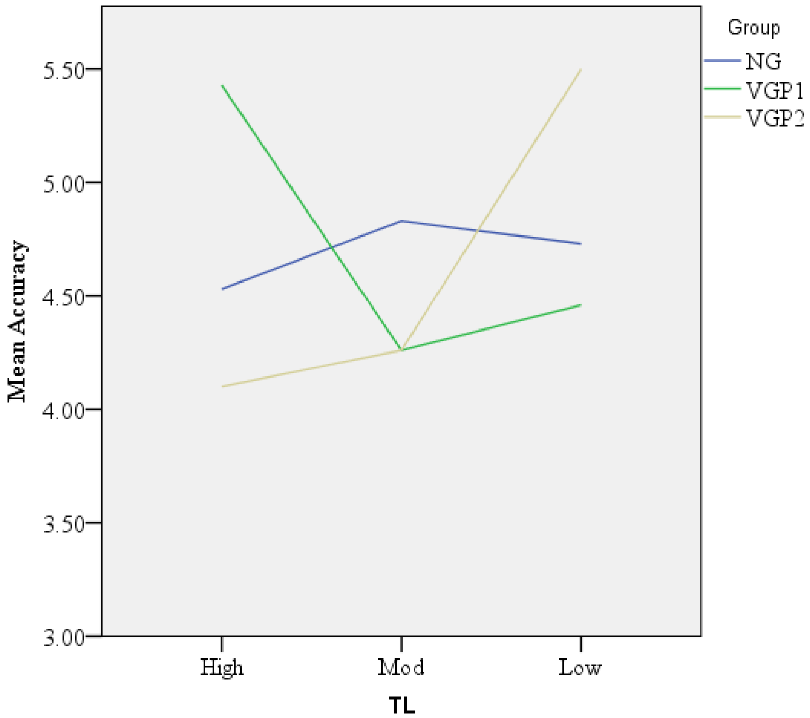


Figure 9. Graph displaying interaction between group and TL in mean accuracy scores across DC.

In sum, participants were significantly more accurate when they were faced with highly critical decisions. However, contrary to hypothesis HP7-A and HP8-A, neither TL or group impacted on the accuracy of decisions.

Table 20: Means and Standard Deviations for Accuracy as influenced by TL, DC, and group

TL	Group	Overall	High DC	Medium DC	Low DC
High					
	NG	13.60 (3.84)	4.80 (1.48)	4.80 (1.87)	4.00 (1.76)
	VGP1	16.30 (3.37)	6.30 (0.82)	5.70 (2.11)	4.30 (1.83)
	VGP2	12.30 (3.53)	5.30 (1.83)	3.80 (1.99)	3.20 (1.22)
	Total	14.07 (3.85)	5.47 (21.52)	4.77 (2.08)	3.83 (1.62)

Moderate	NG	14.50 (3.75)	5.20 (1.81)	4.80 (1.87)	4.50 (2.46)
	VGP1	12.50 (4.01)	4.90 (1.37)	4.90 (1.45)	3.00 (2.11)
	VGP2	12.80 (3.91)	4.50 (0.50)	5.10 (2.33)	3.20 (1.87)
	Total	13.27 (3.86)	4.87 (1.38)	4.93 (1.86)	3.57 (2.22)
Low	NG	14.30 (2.95)	5.80 (2.75)	5.10 (1.66)	3.30 (1.16)
	VGP1	13.40 (1.78)	5.90 (1.10)	4.70 (1.49)	2.80 (1.13)
	VGP2	16.50 (2.84)	7.00 (1.41)	5.80 (1.98)	3.70 (1.16)
	Total	14.73 (2.82)	6.23 (1.50)	5.20 (1.73)	3.25 (1.17)
	NG	14.13 (3.43)	5.27 (1.68)	4.90 (1.75)	3.93 (1.87)
	VGP1	14.07 (3.49)	5.70 (1.24)	5.10 (1.71)	3.37 (1.85)
	VGP2	13.87 (3.84)	5.60 (1.73)	4.90 (2.20)	3.37 (1.43)
Total		14.02 (3.55)	5.52 (1.56)	4.97 (1.88)	3.56 (1.73)

Note. Standard deviations are in parenthesis.

6.4.2. Confidence:

Participants were asked to rate confidence in their decision ‘0’ being *not confident at all* and ‘5’ *being extremely confident*. The maximum confidence score in total was 150 and for each DC 50. A 3 X 3 X 3 mixed ANOVA was carried out to assess the impact of TL, group, and DC on decision confidence (see Table 21).

A main effect of confidence in DC was found F(2,162) = 20.98, $p < .001$, $\eta_p^2 = .21$. A Bonferroni corrected post-hoc test showed that participants were significantly more confident in high DC decisions (*M* = 37.83, *SD* = 8.22) than medium DC decisions (*M* = 34.74, *SD* = 7.95), $p < .001$. Participants were also more confident in high DC compared to Low DC (*M* = 36.52, *SD* = 4.31), $p = .025$ and more confident in low DC that Medium and Low DC $p < .001$. No significant differences were found between high DC and medium DC, $p > .05$. Hence, medium DC decisions reduced individuals' decision confidence. However, no main effect of the TL conditions was found, F(2, 81) = .602, $p = .550$, $\eta_p^2 = .02$. Hence, decision confidence was not influence by the TL conditions.

A significant main effect of confidence was found between groups F(2, 81) = 6.84, $p = .002$, $\eta_p^2 = .15$. VGP2 were significantly more confident (*M* = 39.23, *SD* = 4.94) than NG (*M* = 32.53, *SD* = 7.22), $p = .001$. No significant differences between NG and VGP1, $p = .103$, or VGP1 and VGP2, $p = 391$. An interaction effect was observed between DC and TL F(4, 162) = 3.02, $p = .020$, $\eta_p^2 = .07$.

No significant interaction effects between DC and group F(4, 162) = 1.22, $p = .306$, $\eta_p^2 = .03$. No significant interaction effects between DC, TL and group F(8, 162) = .43, $p = .902$, $\eta_p^2 = .02$. No interaction was observed between TL and group F(4, 81) = .86, $p = .495$, $\eta_p^2 = .04$. The results demonstrate that confidence in high DC decisions reduces significantly in conditions of moderate TL compared to the low TL. Similarly, there is a large reduction in confidence in medium DC from the high to the low TL conditions (see Figure 10).

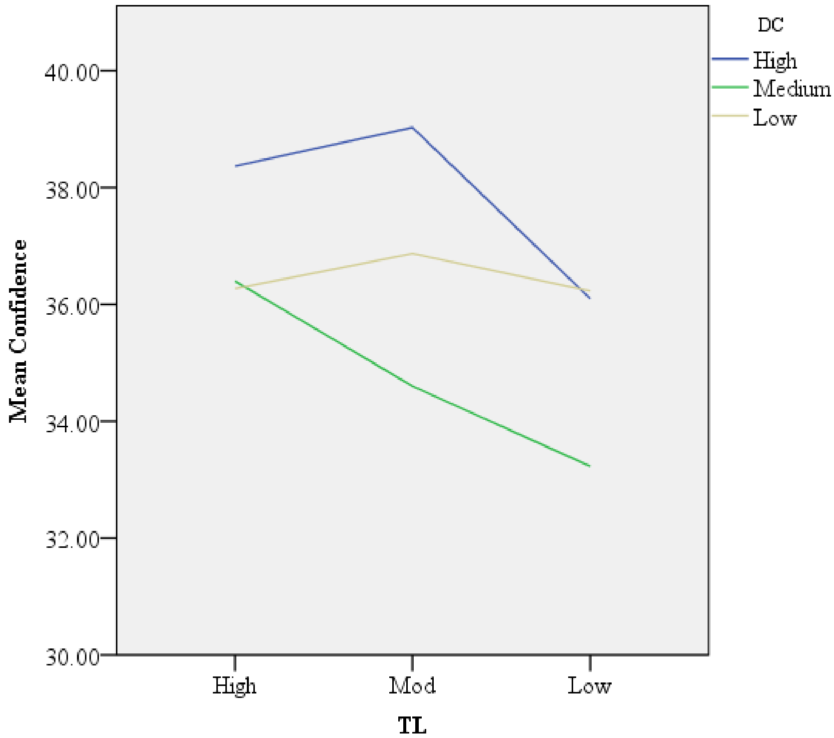


Figure 10. Graph displaying interaction between DC and TL in mean confidence scores.

Table 21: Means and Standard Deviations for Confidence as influenced by DC, TL and group

TL	Group	Overall	High DC	Medium DC	Low DC
High					
	NG	97.30 (23.67)	33.60 (7.49)	32.40 (8.04)	31.40 (9.36)
	VGP1	119.50 (12.87)	40.70 (6.58)	38.30 (8.34)	38.50 (6.85)
	VGP2	118.30 (13.65)	40.90 (5.63)	38.50 (4.53)	38.90 (5.22)
	Total	111.70 (19.76)	38.37 (7.24)	36.40 (7.50)	36.27 (7.90)
Moderate					
	NG	95.60 (18.05)	34.70 (7.70)	30.10 (6.22)	32.30 (6.04)
	VGP1	108.70 (38.73)	38.40 (13.25)	33.30 (12.78)	37.00 (13.08)
	VGP2	125.90	44.00	40.40	41.60

		(11.62)	(3.23)	(4.65)	(4.97)
	Total	110.07 (27.71)	39.03 (9.55)	34.60 (9.41)	36.87 (9.33)
Low					
	NG	98.12 (20.39)	34.10 (8.75)	32.40 (8.26)	34.60 (5.48)
	VGP1	110.83 (25.10)	35.90 (8.08)	31.40 (6.45)	37.20 (3.05)
	VGP2	118.53 (13.73)	38.30 (6.26)	35.90 (4.33)	37.20 (3.97)
	Total	109.17 (21.76)	36.10 (7.69)	33.23 (6.62)	36.23 (4.32)
	NG	100.83 (24.10)	34.10 (8.75)	31.63 (7.38)	32.77 (7.05)
	VGP1	113.07 (18.3)	35.90 (8.08)	34.33 (9.69)	37.57 (8.43)
	VGP2	121.00 (15.6)	38.30 (6.26)	38.27 (4.73)	39.23 (4.94)
Total		111.63 (21.1)	36.10 (7.69)	34.74 (7.95)	36.85 (7.41)

Note. Standard deviations are in parenthesis.

6.4.3. W-S C-A:

A 3 X 3 X 3 mixed ANOVA was performed on the relationship between DC, TL and group on individuals W-S C-A. There was a significant main effect of DC shown, $F(2, 162) = 3.16$, $p = .045$, $\eta_p^2 = .04$. The post hoc comparisons displayed a close to significant difference. The results showed that participants had higher W-S C-A scores in high DC ($M = .11$, $SD = .34$) decisions than low DC decisions ($M = -.02$, $SD = .36$) $p = .067$. Participants were moderately better able to discriminate between correct and incorrect responses when the decisions were highly critical i.e. confident in correct and not confident in incorrect (see Table 22).

No main effect of TL on W-S C-A scores F(2, 81) = .50, *p* = .606, η_p^2 = .01 was found. In addition, no main effect of group F(2, 81) = 1.09, *p* = .340, η_p^2 = .03. Further, no interaction was observed between DC and TL condition F(4, 162) = .82, *p* = .513, η_p^2 = .02. No interaction was observed between DC and group F(4, 162) = .48, *p* = .751, η_p^2 = .01. No interaction between DC, TL and group F(8, 162) = .154, *p* = .214, η_p^2 = .01. There was also no interaction between TL and group F(4, 81) = .05, *p* = .994, η_p^2 = .00. Overall W-S C-A scores were very low and not negative (*M* = .04, *SD* = .22). The findings did support the HP6-C that DC would impact on W-S C-A however, they do not support the HP7-C, that TL would impact on W-S C-A.

Table 22:Means and Standard Deviations for W-S C-A as influenced by DC, TL and group

TL	Group	Overall	High DC	Medium DC	Low DC
High					
	NG	.08 (.24)	.17 (.44)	-.01 (.33)	.01 (.27)
	VGP1	-.05 (.14)	.04 (.34)	-.09 (.35)	-.00 (.46)
	VGP2	-.03 (.22)	.01 (.36)	.01 (31)	-.02 (.36)
	Total	-.00 (.20)	.07 (.38)	-.06 (.32)	-.00 (.37)
Moderate					
	NG	.05 (.20)	.12 (.33)	.00 (.32)	.11 (.28)
	VGP1	-.02 (.02)	.04 (.35)	-.09 (37)	.04 (.37)
	VGP2	.12 (.34)	.20 (.21)	.09 (.31)	-.18 (.28)
	Total	.06 (.26)	.12 (.30)	-.00 (.33)	-.01 (.32)
Low					
	NG	.11 (.25)	.19 (.36)	.17 (.27)	.03 (.28)

	VGP1	.01 (.17)	.18 (.24)	.08 (.31)	-.20 (.42)
	VGP2	.00 (.16)	-.00 (.34)	.09 (.31)	.05 (.52)
	Total	.04 (.20)	.12 (.35)	.11 (.32)	-.04 (.42)
Total					
	NG	.09 (.37)	.16 (.22)	.02 (.32)	.05 (.27)
	VGP1	-.02 (.18)	.09 (.34)	-.04 (.34)	-.05 (.42)
	VGP2	.03 (.25)	.07 (.31)	.06 (.34)	-.05 (.39)
Total		.03 (.22)	.11 (.34)	.02 (.33)	-.02 (.36)

Note. Standard deviations are in parenthesis.

6.4.4. Percentage Confidence in Correct and Incorrect Responses:

To consider the variation in the data and examine the low correlations displayed for W-S C-A, percentage confidence in correct and incorrect responses was calculated. W-S C-A demonstrates the relationship between confidence and accuracy; however, a high correlation suggests both being highly confident in correct decisions as well as low confidence in incorrect decisions. Similarly, a negative correlation would suggest that individuals are highly confident in incorrect responses or not confident in correct responses. Thus, by examining the percentage confidence in incorrect or correct responses the direction of the confidence (over/under confidence) can be displayed (see Table 23).

To do this the number of correct responses was recorded and the confidence in those decisions calculated to produce percentage confidence in correct responses. The same method was applied to incorrect responses.

Interestingly, all data suggests a high degree of confidence in decisions (*M* = 72.60, *SD* = 14.76). Descriptives show that means in correct and incorrect to be the same, percentage correct (*M* = 72.70, *SD* = 14.47) and percentage incorrect (*M* = 72.70, *SD* = 12.47).

Percentage Accuracy in correct responses. To examine the differences in TL and group a two way ANOVA was conducted on percentage confidence in correct responses. No main effect of TL on percentage confidence in correct decisions was observed F(2, 81) = .28, *p* = .754, η_p^2 = .01. A main effect was found on group F(2,81) = 5.02, *p* = .009, η_p^2 = .11. VGP2 (*M* = 78.45, *SD* = 9.05) were significantly more confident in correct responses than NG (*M* = 67.00, *SD* = 14.99), *p* = .002. No differences between NG and VGP 1(*M* = 72.63, *SD* = 16.37), *p* = .123 nor VGP1 and VGP2, *p* =.111. No interaction between group and TL F(4,81) = .83, *p* = .511, η_p^2 = .04 was observed

Percentage Accuracy in incorrect responses. Similary, an ANOVA was conducted with percentatge confidence in incorrect responses. Levene's found to be significant. Again, no main effect percentage incorrect was found in TL F(2, 81) = .35, *p* = 704, η_p^2 = .01. However, a main effect of group was found to be significant F(2, 81) = 7.50, *p* = .001, η_p^2 = .16. Differences were observed between NGs (*M* = 64.60, *SD* =12.92) and VGP2 (*M* = 78.83, *SD* = 8.91) , *p* = .001. No differences between NGs and VGP1 (*M* = 73.13, *SD* = 18.86) *p* = .071 or VGP1 and VGP2, *p* = .381. Furtheremore, no interaction was observed F(4, 81) = .83, *p* = .502, η_p^2 = .04.

These results demonstrate that VGP2s were both more confident in correct and incorrect responses. Individuals in the different TL conditions displayed similar levels of confidence in correct and incorrect responses. This would suggest that confidence is not influenced by the situational demands of the task.

Table 23: Means and Standard Deviations for % confidence (correct and incorrect) according to TL

TL	Group	% Correct	% Incorrect
High	NG	66.30 (15.07)	62.90 (16.39)
	VGP1	77.20 (12.29)	77.80 (15.44)
	VGP2	77.80 (7.71)	78.90 (10.50)
	Total	73.76 (12.83)	73.20 (15.71)
Mod	NG	65.60 (13.23)	63.60 (12.69)
	VGP1	70.80 (24.52)	72.30 (26.91)
	VGP2	83.10 (7.19)	83.00 (8.01)
	Total	73.17 (17.68)	72.97 (18.97)
Low	NG	69.10 (17.72)	67.30 (9.79)
	VGP1	69.90 (8.75)	69.30 (11.91)
	VGP2	74.50 (10.52)	74.60 (6.48)
	Total	71.17 (12.70)	70.40 (9.83)
Total	NG	67.00 (14.99)	64.60 (12.92)
	VGP1	72.63 (16.37)	73.13 (18.86)
	VGP2	78.47 (9.05)	78.83 (8.91)
	Total	72.70 (14.47)	72.19 (15.19)

Note. Standard deviations are in parenthesis.

6.4.5. WL and SA:

Manipulation check: A 3 X 3 ANOVA was conducted to assess the relationship between SA and TL. There was no significant effect of TL on SA F(2, 81) = 2.65, p = .077, η_p^2 = .06. A significant main effect of group was found on SA, F(2, 81) = 6.20, p = .003, η_p^2 = .13. VGP2 players reported having higher SA (M = 21.13, SD = 5.84) than NG (M = 15.60, SD = 6.97), p = .001 and VGP1, p = .038. No differences between NG and VGP1, p = .168. No interaction was found F(4, 81) = .638, p = .637, η_p^2 = .03. This finding suggests that the manipulation of TL did not produce differences in SA however, group differences were observed.

Manipulation check: No significant relationship was found between WL and TL F(2, 81) = 1.85, p = .164, η_p^2 = .04. As a non-significant relationship was found this would suggest that the manipulation check was not successful. A significant relationship was found in groups F(2,81) = 3.76, p = .027, η_p^2 = .09. Post hoc test revealed that NG reported higher levels of WL (M = 64.81, SD = 15.50) than both VGP1 (M = 56.08, SD = 14.35), p = .003 and VGP2 (M = 54.90, SD = 15.79), p = .014. No significant differences between VGP1 and VGP2, p = .766. No interaction between group and TL condition was found F(4,81) = .39, p = .814, η_p^2 = .02. WL was influenced by group, individuals in the NG group reported higher levels of WL. The results from SA and WL show that playing video games impacts on reported feeling of WL and SA after the task.

Table 24: Means and Standard Deviations for WL and SA as influenced by TL and group

TL	Group	Overall SA	Overall WL
High			
	NG	12.99 (9.37)	69.14 (16.92)
	VGP1	15.90 (6.05)	56.27 (12.40)
	VGP2	20.40 (3.83)	59.41 (12.86)

	Total	16.43 (7.27)	61.61 (14.80)
Moderate			
	NG	16.90 (18.42)	63.33 (18.42)
	VGP1	19.30 (6.50)	60.09 (12.93)
	VGP2	24.00 (6.93)	55.23 (15.90)
	Total	20.07 (6.72)	59.85 (16.04)
Low			
	NG	16.90 (5.16)	61.96 (10.59)
	VGP1	18.20 (4.37)	50.99 (16.12)
	VGP2	19.00 (5.70)	50.06 (18.33)
	Total	18.03 (5.15)	54.33 (15.81)
	NG	15.60 (6.97)	64.81 (15.50)
	VGP1	17.80 (5.70)	56.08 (14.35)
	VGP2	21.13 (5.84)	54.90 (15.79)
	Total	18.18 (6.54)	58.59 (15.69)

Note. Standard deviations are in parenthesis.

6.4.6. SA Subscales:

To examine the three dimensions of SA as measured by SART (Demand, Supply, and Understanding) one-way ANOVAs were carried out across each dimension and TL condition (see Table 25).

6.4.6.1. Demand: Attentional demand includes measures which assess individual feelings of the instability of the situation, variability of the situation and complexity of the situation. No significant differences were found for demand of the task for the different task conditions $F(2, 81) = .26$, $p = .774$, $\eta_p^2 = .01$. No effect was observed on group $F(2, 81) = .34$, $p = .712$, $\eta_p^2 = .01$ and no interaction effect $F(4, 81) = .52$, $p = .720$, $\eta_p^2 = .03$.

6.4.6.2. Supply: Attentional supply includes constructs of arousal, spare mental capacity, concentration and division of attention. No significant difference was also found for attentional supply $F(2, 81) = 1.27$, $p = .287$, $\eta_p^2 = .03$. Further, no effect of group $F(2, 81) = 1.49$, $p = .23$, $\eta_p^2 = .04$ and no interaction $F(4, 81) = .38$, $p = .825$, $\eta_p^2 = .02$.

6.4.6.3. Understanding: Understanding measures information quantity, quality and familiarity. Significant differences for understanding were found $F(2, 81) = 4.80$, $p = .011$, $\eta_p^2 = .11$. Comparisons show that those in the high TL rated a lower understanding ($M = 11.53$, $SD = 4.26$) than moderate TL ($M = 14.30$, $SD = 4.25$), $p = .008$. No differences between high and low or low and moderate TL. Also, a significant effect on group was found $F(2, 81) = 7.60$, $p = .001$, $\eta_p^2 = .16$. NG had lower understanding ($M = 11.03$, $SD = 4.15$) than VGP1 ($M = 13.40$, $SD = 3.49$), $p = .010$ and VGP2 ($M = 14.43$, $SD = 3.37$), $p < .001$. However, there was no interaction $F(4, 81) = 1.99$, $p = .104$, $\eta_p^2 = .09$. In sum, these findings demonstrate that the reported level of understanding varied in this task. Understanding was lowest in the high TL and highest in the moderate TL.

Hence, participants rated the amount of knowledge received and understood, the degree of value of knowledge communicated, and the degree of acquaintance with the situation as different across the TL conditions to be different in these conditions. Further, NG reported lower levels of understanding than the gamers. This could be explained by the lack of familiarity to computer-based gaming tasks in comparison to gamers.

Table 25: Means and Standard Deviations for dimensions of SA as influenced by TL and group

TL	Group	Demand	Supply	Understanding
High	NG	12.80 (3.76)	17.00 (4.94)	8.70 (4.50)
	VGP1	14.70 (2.83)	18.30 (4.69)	12.30 (3.43)
	VGP2	13.70 (3.02)	20.50 (2.50)	13.60 (3.50)
	Total	13.73 (3.21)	18.60 (4.30)	11.53 (4.26)
Mod	NG	14.60 (4.35)	20.00 (5.45)	11.40 (4.17)
	VGP1	14.00 (3.71)	19.40 (5.25)	14.70 (4.19)
	VGP2	13.60 (2.72)	20.70 (5.31)	16.80 (2.53)
	Total	14.07 (3.55)	20.03 (5.18)	14.30 (4.25)
Low	NG	14.20 (3.05)	18.10 (2.38)	13.00 (2.71)
	VGP1	13.40 (4.77)	18.40 (2.12)	13.20 (2.57)
	VGP2	12.60 (3.81)	19.10 (2.13)	12.90 (2.88)
	Total	13.40 (3.71)	18.53 (2.18)	13.03 (2.63)
Total	NG	13.87 (3.71)	18.37 (4.49)	11.03 (4.15)
	VGP1	14.03 (3.76)	18.70 (4.13)	13.40 (3.49)
	VGP2	13.30 (3.08)	20.10 (3.56)	14.43 (3.37)
	Total	13.73 (3.50)	19.06 (4.10)	12.66 (3.91)

Note. Standard deviations are in parenthesis

6.4.7. NASA TLX Subscales: As an exploratory analysis, the 6 dimensions of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort and frustration) were also examined to determine whether differences existed across the conditions. One way ANOVAs were conducted with TL and group across the different dimensions of workload (see Table 26).

6.4.7.1 Temporal Demand: There was a significant difference found for TL in temporal demand F(2, 81) = 3.19, p = .046, η_p^2 = .07. Comparisons show a significant difference between the conditions high (M = 257.00, SD = 23.98) and low (M = 175.83, SD = 23.98), p = .019. No significant difference between high and moderate, p = .621 or low and moderate p =.061. The findings suggest that participants felt more time pressure due to the rate and pace at which the task elements occurred in the high TL conditions in comparison to the low task condition.

6.4.7.2. Performance: There was a significant main effect of group on performance F(2,81) = 6.66, p = .002, η_p^2 = .14. NG reported significantly higher levels of performance (M = 178.33, SD = 96.43) than both VGP1 (M = 126.67, SD = 66.14), p = .015, and VGP2s (M = 104.50, SD = 76.37), p = .001. No main effect of TL F(2,89) = .28, p = .760, η_p^2 = .14 and no interaction was observed F(4, 81) = 1.55, p = .196, η_p^2 = .71.The findings therefore suggest that the main difference between the TL conditions was the speed at which the task events occurred. NG also reported higher levels of performance than gamers.

6.4.7.3. Mental Demand, Effort, Physical Demand, Frustration: The following was observed. No main effects of group was found for Mental Demand F(2, 81) = .53, p = .948, η_p^2 = .00. No main effect of TL F(2, 81) = 2.40, p = .097, η_p^2 = .06. No interaction F(4, 81) = 1.72, p = .154, η_p^2 = .08. No main effects on group for physical Demand F(2, 81) = 1.92, p = .152, η_p^2 = .05. No main effect of TL F(2, 81) = .71, p = .493, η_p^2 = .02. No interaction F(4, 89) = .53, p = .712, η_p^2 = .03. No main effect on group on effort F(2, 81) = .84, p = .436, η_p^2 = .02. In addition no main effect of TL on effort was found F(2,81) = .05, p = .954, η_p^2 = .00. No interaction

F(4,81) = .68, p = .604, η_p^2 = .03. No main effect of group on frustration F(2,81) = 2.13, p = .126, η_p^2 = .05 or a main effect of TL F(2,81) = 1.63, p = .202, η_p^2 = .04. No interaction between group and TL F(4,81) =.70, p = .596, η_p^2 = .03.

Table 26: Means and Standard Deviations for dimensions of WL as influenced by TL and group

TL	Group	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
High	NG	299.00 (99.41)	1.00 (3.12)	282.00 (135.12)	212.00 (100.24)	105.50 (74.07)	138.00 (117.24)
	VGP1	301.00 (89.84)	5.50 (11.41)	234.00 (91.65)	115.00 (35.28)	143.00 (91.99)	46.00 (45.02)
	VGP2	326.50 (52.87)	7.00 (14.94)	255.00 (138.66)	76.50 (50.67)	157.00 (99.59)	69.50 (67.02)
	Total	308.83 (81.25)	4.50 (10.94)	257.00 (120.99)	134.50 (87.58)	135.17 (88.86)	84.50 (88.68)
Mod	NG	194.00 (148.47)	.00 (.00)	293.50 (139.27)	166.50 (105.94)	119.00 (79.19)	177.50 (164.57)
	VGP1	276.50 (142.73)	11.00 (28.46)	206.00 (129.93)	129.50 (57.95)	150.50 (110.39)	139.00 (131.19)
	VGP2	280.50 (67.39)	3.00 (7.89)	221.00 (105.80)	94.00 (59.53)	141.00 (109.94)	89.50 (122.10)
	Total	250.33 (127.35)	4.67 (17.12)	240.17 (127.45)	130.00 (80.81)	136.83 (98.28)	135.33 (143.73)
Low	NG	303.50 (114.55)	.00 (.00)	215.50 (154.69)	156.50 (82.02)	135.00 (67.37)	120.40 (87.77)
	VGP1	244.50 (136.09)	3.00 (9.48)	131.50 (147.22)	135.50 (96.19)	155.50 (85.32)	111.00 (98.51)
	VGP2	214.50 (145.34)	1.00 (3.16)	180.50 (127.79)	143.00 (100.20)	99.50 (77.51)	115.00 (93.57)
	Total	254.17 (133.38)	1.33 (5.71)	175.83 (143.00)	145.00 (90.30)	130.00 (78.01)	115.47 (90.20)

Total NG	265.50 (128.95)	.33 (1.83)	263.67 (142.60)	178.33 (96.43)	119.83 (67.37)	145.30 (125.11)
VGP1	274.00 (122.10)	6.50 (18.20)	190.50 (128.49)	126.67 (66.15)	149.67 (93.24)	98.67 (102.72)
VGP2	273.83 (104.97)	3.67 (9.91)	218.83 (124.37)	104.50 (76.37)	132.50 (96.44)	91.33 (100.20)
Total	271.11 (118.13)	3.50 (12.14)	224.33 (134.02)	136.50 (85.58)	134.00 (87.82)	111.77 (111.30)

Note. Standard deviations are in parenthesis.

6.4.8. Between Subject Confidence - Accuracy:

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was also conducted. No significant relationship was found between-subjects confidence and accuracy, $r(88) = -.075$, $p = .485$. Corrections were applied for multiple comparison correlations; a new alpha level of $p = .005$ was used.

6.4.9. Relationship between WL, SA and Accuracy, Confidence and W-S C-A:

To establish whether a relationship existed between WL, SA, accuracy, confidence and W-S C-A a number of Pearson's correlations were carried out.

Results revealed a significant negative relationship between overall WL and confidence, $r(88) = -.392$, $p < .001$. As subjective measures of WL increased, confidence in decisions decreased. In addition, a significant strong positive relationship was found between overall SA and confidence, $r(88) = .504$, $p < .001$. As such, higher scores in subjective SA were related to higher scores of confidence in decisions. However, no significant relationships were found between SA, WL and W-S C-A, or between SA and accuracy or WL and accuracy in decisions; all comparisons, $p > .005$. The findings suggest that decision confidence influences both WL and SA. In this study, accuracy was found to be unrelated to WL and SA.

To assess the relationship between WL and SA, a series of Pearson's correlations were calculated. A significant negative relationship was found between SA and WL, *r(88)* = -.37, $p < .001$. Higher levels of reported WL were related to lower feelings of SA. No other analysis found to be significant $p > .005$.

6.4.10. Personality Constructs:

This study was also interested in establishing whether accuracy, confidence, and W-S C-A were related to the psychometric scores. For this, Pearson's correlations were conducted. Due to the multiple correlation correction, these were found not to be significant. Close to significant results are reported. Accuracy was significantly related to the personality trait of Openness to experience $r(88) = .26$, $p = .013$. Confidence was found to be significantly negatively related to Agreeableness $r(88) = -.21$, $p = .043$, such that those with higher levels of agreeableness reported lower levels of decision confidence. No other measures were found to be related to confidence, $p > .005$.

Ambiguity was negatively related to accuracy, $r(88) = -.23$, $p =.029$. A negative relationship was also found between Decision Style and Accuracy $r(88) = -.22$, $p = .038$. High scorers on the decision style scale were less accurate. Decision Style explicitly probes the need for quick and unambiguous answers. To investigate whether there were individual differences in participants experiences of WL and SA correlations were conducted on each measure of NEO-PI-R (openness to experience, conscientiousness, extraversion, agreeableness and neuroticism) SA was closely related to extraversion $r(88) = .25$, $p = .018$ and conscientiousness $r(88) = .25$, $p = .017$. No other relationships were found to be significant, $p > .005$. These finding, therefore, suggest that some cognitive constructs are involved in decision accuracy and individuals differences in participants' feelings of SA.

To assess and compare each group on the psychometric measures, one-way ANOVAs were performed on the data. No significant differences found between the groups and personality traits. Due to the lack of significant findings, the results table reporting the means and standard deviation of psychometric scores as influenced by group are reported in Appendix 10b.

6.5. Discussion

Following on from the first experiment, the second experiment in the thesis examined how time pressure may influence the variables decision accuracy, confidence and metacognition. In this experiment, the time to make a decision was reduced from the previous experiment from 20 seconds to 10 seconds. The results from this experiment largely mirrored some of those found in Experiment 1. As such, this experiment found that DC impacted on decision confidence and accuracy in the same way in the first experiment. However, the task manipulation of WL was not found to be successful and no differences were found in WL between the TL conditions. Other findings from the experiment include an interaction between group and TL in decision accuracy as well as demonstrating individual differences in decision making.

6.5.1. Accuracy:

Supporting the findings of Experiment 1 and HP6- A, individuals were more accurate in high DC than medium or low DC. Thus supporting the findings that criticality has a positive impact on performance (Harris, Hancock, Arthur & Caird, 1995). However, contrasting the first experiment and, contrary to HP7-A, this study found no differences in the accuracy of decisions between TL conditions. Hence, although participants in this experiment had less time to make a decision than the first experiment, this was not reflected by any significant differences in the accuracy of decisions between conditions. An explanation for this may arise from the manipulation check of WL being unsuccessful as demonstrated by no significant differences between the TL conditions.

With regards to HP9-A, supporting the results from Experiment 1, no differences were found between groups in the accuracy of decisions.

However, there was a significant interaction between group and TL. As such, VGP1 made more accurate decisions in high TL in comparison, VGP2 made more accurate decisions in the low TL. The accuracy of decisions by NGs was not influenced by TL. There is a lack of research into the amount of game play and the influence on decision making. Nevertheless, this interaction suggests that TL interacts with the amount of time playing video games. Hence, the more time spent playing games leads to a better performance when the task is less cognitively demanding.

6.5.2. Confidence:

HP6-B stated that high DC would reduce decision confidence. However, contrary to this, individuals were more confident in both high DC decisions and low DC decisions. Medium DC decisions produced the most uncertainty in decisions, as expressed by lower decision confidence. Thus, again supporting the findings from the first experiment and that the medium decision events reduce confidence by their increased uncertainty. Conflicting HP7-B, but in support of the first experiment, decision confidence was not found to be influenced by TL. Similar levels of decision confidence were provided for each TL condition. This second experiment also found support for HP9-B, with VGP2s more confident in their decisions which again provide further support for the finding of the first experiment.

In addition, the results from this experiment found an interaction between DC and TL. Individuals tended to be more confident in high DC decisions when in a low TL condition compared to when they were in a moderate TL. Low TL also reduced confidence in medium DC decisions. This interaction could suggest that a reduced temporal and cognitive load i.e. low TL, allowed individuals to be more confident in their high DC events, in comparison, confidence was hindered in high DC events in the moderate load conditions. As such, the increased load hindered confidence in highly critical decisions. Hence, the condition in which the highest confidence was found was in a low TL, in high DC decisions.

6.5.3. W-S C-A:

Unlike the first experiment, significant differences were found for W-S C-A for the different decision criticalities. The findings demonstrated that individuals had a better W–S C-A relationship when they were making highly critical decisions compared to the low criticality decisions. As such, participants were better able to discriminate between correct and incorrect responses when the decisions were highly critical, i.e., confident in correct and not confident in incorrect. The findings would, therefore, suggest that individuals seem more able to apply greater sensitivity when decisions held more consequences. Hence, metacognitive ability in this task was improved for highly critical decisions. An explanation for this finding is that individuals may have been more aware of the importance of these decisions (Kliegel, Martin, McDaniel & Einstein, 2004). Crucially, this finding demonstrates that criticality does not impair metacognition. However, the overall calibration was low and analysis of percentage confidence would suggest that there is a tendency for overconfidence across all DC and TL. Nevertheless, such a finding is an important one.

6.5.4. WL and SA:

Contrasting the first experiment, there were no differences in global SA or global WL between the TL conditions. Hence, in the second experiment, the TL manipulation was not significant. However, temporal demand in WL was significantly higher in the high TL condition. This suggests that although global WL was not influenced by TL, individuals did report a difference in the speed of the task between the TL conditions. This could be due to the reduction in time to make a decision as well as the increased frequency of the decision in this condition. However, the manipulation of the reduction in time to make a decision did not reflect impairments of global WL or any other WL subscale.

In support of the first experiment, Pearson's Correlations revealed that SA was positively related to decision confidence. In comparison, WL was negatively related to confidence. Hence, individuals who reported higher levels of confidence in their decisions also reported higher SA and those who reported higher WL, were less confident in their decisions. In addition, contrary to the first experiment, differences were found between groups in global SA and WL in this experiment. VGP2 reported higher SA than NGs. Hence, the experience of playing games could have led them to believe that they had a better understanding of the task. Research has shown that gaming is related to SA (Chiappe, Conger, Liao, Cladwell & Vu, 2013). This is contrary to the findings of Experiment 1 and Vidulich et al. (1995). However, as VGPs were also more confident in their decision, an explanation for this finding has been linked to confidence in general and self-reported SA being similar measures (Endsley, 1995). Nevertheless, further work on objective SA would be needed to assess this relationship in more detail.

Differences in subscales of SA revealed that the reported level of understanding varied in this task. Understanding was lowest in the high TL and highest in the moderate TL. Hence, participants rated the amount of knowledge received and understood, the degree of value of knowledge communicated, and the degree of acquaintance with the situation as the different across the TL conditions. More information to process and less time to in between decisions had an impact on understanding. Furthermore, groups also reported differences in understanding. NG reported lower levels of understanding than the gamers. This could be related to the familiarity that gamers may have, and increased WL experienced by NGs.

In regards to WL, contrary to the first experiment which found no differences in WL and gaming experience, NGs reported higher WL compared to gamers. As mentioned, research has found that playing video games increases cognitive ability (Boot et al., 2008; Hubert-Wallander et al., 2011) and high WL is more detrimental in individuals with low cognitive abilities than high (Gonzalez, 2004). The results from this study on feelings of SA and WL show that playing video games does impact on the reported feeling of WL and SA after the task. Subsequently, high SA and low WL would be beneficial in complex environments; therefore, it could be beneficial to investigate VGP2s in highly complexed environments as this group showed to demonstrate this ability.

6.5.5. Personality and Cognitive Constructs:

The cognitive constructs of ambiguity and decision style measured in this experiment were found to be negatively related to making accurate decisions. This finding supports the findings of the first experiment and provided further support that a higher Tolerance to Ambiguity is beneficial in complex and uncertain environments. Additionally, the trends in this study demonstrate that decision accuracy was found to be related to the personality trait of openness to experience. This validates the findings of Schaefer et al. (2004). It has been argued that individuals with higher openness have been found to have higher levels of intellect and get bored of routine (John, Naumann & Soto, 2008) and are therefore better suited to a more dynamic environment (Colbert, Barrick & Bradley et al., 2014). Hence, contrary to the first experiment, these findings demonstrate a potential link between personality and performance.

Furthermore, the trends in this experiment show that decision confidence was found to be negatively related to agreeableness. Individuals high in agreeableness are more sympathetic and trusting, lower confidence. For example, it has been previously argued that individuals with high levels of agreeableness tend to work better in teams (John et al., 2008) and may doubt their own decision leading to more decision bias which may reduce confidence in decisions (Erjavec, Jure & Trkman 2016).

In addition, in relation to SA, high SA was also found to be related to higher levels of extraversion and conscientiousness. Thus, providing further support those traits of high extraversion and conscientiousness would be advantages in these environments. However, again it is important to be mindful of the self-reporting nature of the SA score and the relationship of these factors to confidence/self-efficacy.

6.6. Summary

In summary, this experiment found support for some of the findings displayed in the first experiment. DC impacted on both decision confidence and decision accuracy. On the other hand, TL was not found to impact on accuracy or confidence. Additionally, this experiment found that individuals were better able to discriminate between their accurate and inaccurate decisions when presented with higher critical decisions. Nevertheless, the time pressure created by reducing the time from 20 seconds to 10 seconds to make a decision did not influence global feelings of WL and SA on decision accuracy, confidence and W-S C-A between the TL conditions. It did however; increase feelings of temporal demand on the task which would suggest that it had some effect on individuals feeling of speed during the task. In addition, evidence for individual differences in decision accuracy and confidence was also found. As such, additional support was provided to the findings in the first experiment which found that a tolerance to uncertain situations is advantageous for accurate decision making. The findings also suggest that conscientiousness and extraversion are related to SA. Next, to examine the differences between the first and second experiment in more detail Chapter seven ran a comparison analysis on the two sets of data to examine whether there were any differences generated by the time to make a decision.

CHAPTER SEVEN

7. Decision Time Comparison Analysis: 10 seconds vs 20 seconds

7.1. Introduction

Chapter 7 presents the results from a comparing the data collected from the first two experiments. As previously reported in the first experiment, (Chapter 5) participants had 20 seconds to make a decision compared to the second experiment (Chapter 6) in which participants were given 10 seconds to make a decision. As discussed in Chapter two, time constraints influence how individuals make decisions. However, there is limited information on the time needed to induce feelings of time pressure on a task (Maule & Hockey, 2000). Hence, to establish if there were any differences on time to make a decision, the data from each experiment was collated together and re-analysed. In addition, the findings from this analysis could potentially be used to explore the time period needed in order to assess the levels of time pressure experienced by decision makers in this task. This chapter reports the findings collected from comparing the first and second experiment.

7.2. Hypotheses

- **HP11**. There will be differences between the time to make a decision in decision accuracy, confidence and W-S C-A.
- **HP12**. There will be differences between the time to make a decision in WL and reduce SA.

7.3. Participants

In total, the data from 180 participants was analysed. Ninety participants from each experiment were used. The mean age was 23 years old (*SD* = 5.60).

7.4. Results

To assess the differences in means a number of statistical analyses were performed on the data for accuracy, confidence and W-S C-A using

Analysis of Variance (ANOVA). A manipulation check was carried out to assess the differences in TL (see analysis of Workload and Situational Awareness). An alpha level of .05 was used for all statistical tests.

7.4.1. Accuracy:

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X (Time: 10s, 20s) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data (see Table 27).

A main effect of DC was found on decision accuracy $F(2,324) = 57.56$, $p < .001$, $\eta_p^2 = .26$. Post hoc comparisons showed that participants were more accurate in high DC decisions ($M = 5.38$, $SD = 1.75$) than medium ($M = 4.79$, $SD = 1.80$), $p = .001$ and low DC ($M = 3.58$, $SD = 1.78$), $p < .001$. Participants were also more accurate in medium DC than low DC, $p < .001$. No significant interactions with DC. These findings suggest that individuals were less accurate when presented with a low DC event.

A main effect of TL was also observed on decision accuracy $F(2,162) = 5.87$, $p = .003$, $\eta_p^2 = .07$. Participants were more accurate in low TL ($M = 4.94$, $SD = 1.26$) than moderate TL ($M = 4.22$, $SD = .88$), $p = .002$. No other comparisons were significant $p > .05$. In addition, no differences between groups was observed $F(2,162) = .03$, $p = .972$, $\eta_p^2 = .00$. Hence, playing video games did not impact on decision accuracy.

Importantly, no significant differences in accuracy between the conditions of time condition to make a decision was found $F(1, 162) = .423$, $p = .516$, $\eta_p^2 = .003$. Hence, time to make a decision did not have an impact on the accuracy of individual's decisions.

A significant interaction between group and TL was found $F(4,162) = 4.00$, $p = .004$, $\eta_p^2 = .09$. The findings demonstrate that NG displayed similar accuracy across conditions whereas VGP1 performed better in high TL compared to VGP2 better in low TL. No other interactions were found to be significant $p > .05$.

Table 27: Means and Standard Deviations for Accuracy as influenced by DC, TL, time and group

TL	Group	Time Condition	High DC	Medium DC	Low DC
High	NG	10s	4.80 (1.48)	4.80 (1.87)	4.00 (1.76)
		20s	4.90 (2.56)	4.60 (2.12)	4.20 (1.55)
		Total	4.85 (2.03)	4.70 (1.95)	4.10 (1.62)
	VGP1	10s	6.30 (.82)	5.70 (2.11)	4.30 (1.83)
		20s	5.30 (1.95)	5.00 (1.63)	4.00 (2.71)
		Total	5.80 (1.54)	5.35 (1.87)	4.15 (2.25)
	VGP2	10s	5.30 (1.83)	3.80 (1.99)	3.20 (1.23)
		20s	5.10 (1.66)	3.40 (1.17)	3.80 (1.40)
		Total	5.20 (1.70)	3.60 (1.60)	3.50 (1.32)
	Total	10s	5.47 (1.53)	4.77 (2.08)	3.83 (1.64)
		20s	5.10 (2.02)	4.33 (1.77)	4.00 (1.91)
		Total	5.28 (1.79)	4.55 (1.93)	3.92 (1.77)
Mod	NG	10s	5.10 (1.85)	4.60 (1.84)	3.70 (2.11)
		20s	4.40 (1.78)	4.60 (2.07)	3.60 (2.01)
		Total	4.75 (1.80)	4.60 (1.90)	3.65 (2.01)
	VGP1	10s	4.70 (1.64)	5.00 (1.25)	3.00 (2.21)
		20s	4.60 (1.58)	4.20 (1.87)	3.10 (1.60)

		Total	4.65 (1.57)	4.60 (1.60)	3.05 (1.88)
	VGP2	10s	4.50 (.85)	5.10 (2.33)	3.20 (1.87)
		20s	5.10 (1.10)	4.90 (1.10)	2.60 (2.12)
		Total	4.80 (1.01)	5.00 (1.78)	2.90 (1.97)
	Total	10s	4.77 (1.48)	4.90 (1.81)	3.30 (2.02)
		20s	4.70 (1.49)	4.57 (1.70)	3.10 (1.90)
		Total	4.73 (1.47)	4.73 (1.75)	3.20 (1.95)
Low	NG	10s	5.80 (1.75)	5.10 (1.66)	3.30 (1.16)
		20s	6.20 (2.04)	5.10 (2.13)	3.70 (1.25)
		Total	6.00 (1.86)	5.10 (1.86)	3.50 (1.19)
	VGP1	10s	5.90 (1.10)	4.70 (1.49)	2.80 (1.14)
		20s	5.40 (1.43)	4.40 (1.27)	3.60 (1.90)
		Total	5.65 (1.27)	4.55 (1.36)	3.20 (1.58)
	VGP2	10s	7.00 (1.41)	5.80 (1.99)	3.70 (1.16)
		20s	6.40 (2.32)	5.40 (1.65)	4.60 (2.12)
		Total	6.70 (1.90)	5.60 (1.79)	4.15 (1.73)
	Total	10s	6.23 (1.50)	5.20 (1.73)	3.27 (1.17)
		20s	6.00 (1.95)	4.97 (1.71)	3.97 (1.79)

		Total	6.12 (1.73)	5.08 (1.71)	3.62 (1.54)
Total	NG	10s	5.23 (1.70)	4.83 (1.74)	3.67 (1.69)
		20s	5.17 (2.21)	4.77 (2.05)	3.83 (1.60)
		Total	5.20 (1.96)	4.80 (1.89)	3.75 (1.63)
	VGP1	10s	5.63 (1.38)	5.13 (1.66)	3.37 (1.85)
		20s	5.10 (1.65)	4.53 (1.59)	3.57 (2.08)
		Total	5.37 (1.53)	4.83 (1.64)	3.47 (1.95)
	VGP2	10s	5.60 (1.73)	4.90 (2.20)	3.37 (1.43)
		20s	5.53 (1.81)	4.57 (1.55)	3.67 (2.02)
		Total	5.57 (1.76)	4.73 (1.89)	3.52 (1.74)
	Total	10s	5.49 (1.60)	4.96 (1.87)	3.47 (1.65)
		20s	5.27 (1.90)	4.62 (1.73)	3.69 (1.89)
		Total	5.38 (1.75)	4.79 (1.80)	3.58 (1.78)

Note. Standard deviations are in parenthesis

7.4.2. Confidence:

Participants were asked to rate confidence in their decision ‘0’ being *not confident at all* and ‘5’ *being extremely confident*. The maximum total confidence score was 150 and for each DC 50. A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 2 (Time, 20s, 10s) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data.

Transformations were considered only when more than two levels failed the normality or homogeneity of variance assumptions. Shapiro-Wilk was found to be significant in two levels, however, all other conditions met the tests of normality and ANOVA has been demonstrated to be robust for violations of normality (Schmider, Ziegler, Danay, Beyer & Bühner, 2010). As such, the data was not transformed. Mauchly's test of sphericity was found to be significant as such Greenhouse-Geisser *df* is reported (see Table 28).

A main effect of DC on confidence $F(1.81, 293.88) = 22.53$, $p < .001$, $\eta_p^2 = .12$ was found. Participants were more confident in high DC ($M = 38.31$, $SD = 8.57$) decisions than medium DC ($M = 35.49$, $SD = 7.69$), $p < .001$ as well as low ($M = 37.133$, $SD = 8.21$), $p = .049$. An interaction between DC, TL and Time condition was also found $F(4, 293.88) = 2.63$, $p = .040$, $\eta_p^2 = .03$. Figure 11 displays this interaction. The interaction suggests that in 10 seconds, in the moderate and low TL, individuals were less confident in high, medium and low DC events compared to the high TL condition. In the high TL condition individuals were more confident in the medium and high decisions when given 10 seconds to make a decision. In comparison individuals were more confident in low DC events when they were given 20 seconds to make a decision.

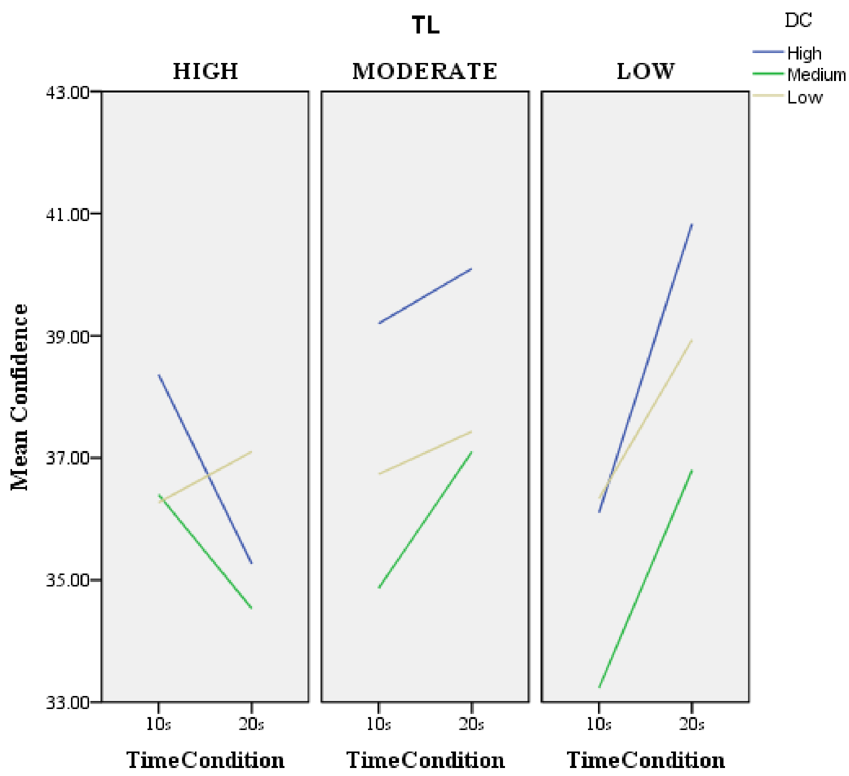


Figure 11. Graph displaying interaction between DC, TL and Time condition in Confidence

Additionally, a main effect of group on confidence was observed F(2, 162) = 14.64, p < .001, η_p^2 = .15. VGP2 (M = 40.44, SD = 1.74) were more confident in their decisions than both NGs (M = 33.48, SD = 1.24), p < .001 and VGP1 (M = 37.01, SD = 1.56), p = .020. VGP2 were also more confidence than VGP1, p = .025. Hence, the results demonstrate that time constraints did not impact on decision confidence. No other interactions found to be significant, p > .05.

Table 28: Means and Standard Deviations for Confidence as influenced by DC, TL, time and group

TL	Group	Time Condition	High DC	Medium DC	Low DC
High	NG	10s	33.60 (7.49)	32.40 (8.04)	31.40 (9.36)
		20s	31.10 (11.48)	30.00 (10.78)	36.20 (18.77)
		Total	32.35 (9.52)	31.20 (9.34)	33.80 (14.65)

	VGP1	10s	40.70 (6.58)	38.30 (8.34)	38.50 (6.85)
		20s	33.00 (12.59)	34.70 (5.93)	36.70 (6.17)
		Total	36.85 (10.55)	36.50 (7.28)	37.60 (6.41)
	VGP2	10s	40.80 (5.63)	38.50 (4.53)	38.90 (5.22)
		20s	41.70 (5.95)	38.90 (6.71)	38.40 (5.48)
		Total	41.25 (5.66)	38.70 (5.57)	38.65 (5.21)
	Total	10s	38.37 (7.24)	36.40 (7.50)	36.27 (7.90)
		20s	35.27 (11.10)	34.53 (8.64)	37.10 (11.47)
		Total	36.82 (9.42)	35.47 (8.08)	36.68 (9.77)
Mod	NG	10s	35.20 (8.43)	30.90 (6.72)	31.60 (6.59)
		20s	35.60 (7.56)	32.90 (7.62)	34.30 (10.85)
		Total	35.40 (7.80)	31.90 (7.07)	32.95 (8.85)
	VGP1	10s	38.40 (13.25)	33.30 (12.78)	37.00 (13.08)
		20s	40.20 (7.25)	37.20 (6.11)	37.80 (6.53)
		Total	39.30 (10.44)	35.25 (9.95)	37.40 (10.07)
	VGP2	10s	44.00 (3.23)	40.40 (4.65)	41.60 (4.97)
		20s	44.50 (3.47)	41.20 (2.74)	40.20 (7.12)

		Total	44.25 (3.28)	40.80 (3.74)	40.90 (6.02)
	Total	10s	39.20 (9.67)	34.87 (9.40)	36.73 (9.57)
		20s	40.10 (7.17)	37.10 (6.62)	37.43 (8.46)
		Total	39.65 (8.45)	35.98 (8.14)	37.08 (8.96)
Low	NG	10s	34.10 (8.75)	32.40 (8.26)	34.60 (5.48)
		20s	37.00 (4.97)	33.90 (4.63)	35.50 (4.60)
		Total	35.55 (7.08)	33.15 (6.56)	35.05 (4.95)
	VGP1	10s	35.90 (8.08)	31.40 (6.45)	37.20 (3.05)
		20s	40.30 (7.13)	36.40 (6.11)	39.20 (5.43)
		Total	38.10 (7.75)	33.90 (6.63)	38.20 (4.41)
	VGP2	10s	38.30 (6.26)	35.90 (4.33)	37.20 (3.97)
		20s	45.20 (6.55)	40.10 (8.36)	42.10 (6.19)
		Total	41.75 (7.17)	38.00 (6.83)	39.65 (5.65)
	Total	10s	36.10 (7.69)	33.23 (6.62)	36.33 (4.32)
		20s	40.83 (6.96)	36.80 (6.83)	38.93 (5.93)
		Total	38.47 (7.66)	35.02 (6.91)	37.63 (5.31)
Total	NG	10s	34.30 (7.98)	31.90 (7.47)	32.53 (7.22)

	20s	34.57 (8.54)	32.27 (7.97)	35.33 (12.38)
	Total	34.43 (8.19)	32.08 (7.66)	33.93 (10.15)
VGP1	10s	38.33 (9.60)	34.33 (9.69)	37.57 (8.43)
	20s	37.83 (9.66)	36.10 (5.93)	37.90 (5.94)
	Total	38.08 (9.55)	35.22 (8.02)	37.73 (7.23)
VGP2	10s	41.03 (5.56)	38.27 (4.73)	39.23 (4.94)
	20s	43.80 (5.51)	40.07 (6.24)	40.23 (6.27)
	Total	42.42 (5.66)	39.17 (5.56)	39.73 (5.62)
Total	10s	37.89 (8.28)	34.83 (7.94)	36.44 (7.50)
	20s	38.73 (8.88)	36.14 (7.42)	37.82 (8.85)
	Total	38.31 (8.57)	35.49 (7.69)	37.13 (8.21)

Note. Standard deviations are in parenthesis

7.4.3. W-S C-A:

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 2(Time, 20s, 10s) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data. The analysis found no significant main effects for DC, $F(2, 322) = 2.36$, $p = .096$, $\eta_p^2 = .014$, TL $F(2, 161) = 1.64$, $p = .198$, $\eta_p^2 = .02$, group $F(2, 161) = 1.10$, $p = .335$. $\eta_p^2 = .01$ or time $F(1, 161) = .09$, $p = .765$, $\eta_p^2 = .00$. Furthermore, no interactions with TL or DC were found to be significant $p > .05$. This supports previous findings in Experiment 1 and 2. Due to the lack of significant findings, the results table reporting the means and standard deviation of W-S C-A as influenced by DC and TL are reported in the Appendix 10c.

7.4.4. WL and SA:

To assess whether the differences in time to make a decision impacted on SA, a 3 X 3 X 2 ANOVA was conducted on the data. There was no main effect of time F(2, 162) = .62, p = .43, η_p^2 = .004. There was however a main effect of TL F(2, 162) = 6.24, p = .002, η_p^2 = .07. Higher SA was reported in the low TL (*M* = 19.07, *SD* = 4.64) compared to high TL (*M* = 15.62, *SD* = 6.39), p = .005. SA was also significantly higher in moderate TL (*M* = 18.80, *SD* = 7.04) than high TL p = .012. No differences between low and moderate TL, p = 1.00.

There was also a main effect of group F(2, 162) = 5.54, p = .005, η_p^2 = .06. Post hoc show that VGP2 (*M* = 19.78, *SD* = 5.44) reported higher SA that NG (*M* = 16.22, *SD* = 6.89) (p = .004). No other significant differences were reported, SA, TL or group on W-S C-A and no interactions were observed.

To examine whether time to make a decision had an impact on WL a 3 X 3 X 2 ANOVA was conducted on TL, group and time. No significant differences were found between overall WL and between the conditions of 10 seconds and 20 seconds F(2, 162) = .273, p = .602, η_p^2 = .02. This would suggest that a reduction in the time from 20 seconds to 10 seconds did not increase perceived WL in the task. There was a main effect of TL was observed on WL F(2, 162) = 10.13, $p < .001$, η_p^2 = .11. Post hoc comparisons revealed significant differences between high and low TL. Higher WL reported TL in the high TL condition (*M* = 63.41, *SD* = 12.81) than low TL (*M* = 51.33, *SD* = 15.75), $p < .001$. Differences were also observed between moderate TL (*M* = 59.30, *SD* = 15.58) and low TL, p = .012. No differences were observed between high and moderate TL, p = .400. A main effect of group on WL was observed, F(2, 162) = 5.10, p = .01, η_p^2 = .06. Post hoc comparisons revealed that NG (*M* = 63.04, *SD* = 15.28) reported higher levels of WL than VGP1 (*M* = 55.70, *SD* = 14.80), p = .024 and VGP2 (*M* = 55.30, *SD* = 16.34), p = .015. No differences between VGP1 and VGP2, *p* = 1.00, and no interaction effects were observed. In sum, WL was influenced by both group and TL. VGPs reported less WL and the high TL increased feelings of WL. Time to make a decision did not influence feelings of WL.

Table 29: Means and Standard Deviations for WL and SA as influenced by TL, time and group

TL	Group	Time Condition	WL	SA
High	NG	10s	69.14 (16.92)	12.99 (9.37)
		20s	65.28 (14.05)	13.70 (6.06)
		Total	67.21 (15.26)	13.35 (7.69)
	VGP1	10s	56.27 (12.40)	15.90 (6.05)
		20s	62.90 (16.23)	13.60 (4.22)
		Total	59.58 (12.46)	14.75 (5.21)
	VGP2	10s	59.41 (12.56)	20.40 (3.84)
		20s	67.48 (7.16)	17.10 (5.45)
		Total	63.45 (10.94)	18.75 (4.89)
	Total	10s	61.61 (14.80)	16.43 (7.27)
		20s	65.22 (12.75)	14.80 (5.37)
		Total	63.41 (12.81)	15.62 (6.39)
Mod	NG	10s	63.33 (18.52)	16.90 (5.11)
		20s	63.50 (16.64)	16.90 (8.89)
		Total	63.42 (17.14)	16.90 (7.06)

	VGP1	10s	60.98 (12.93)	19.30 (6.50)
		20s	56.94 (14.82)	18.20 (8.22)
		Total	58.96 (14.15)	18.75 (7.23)
	VGP2	10s	55.23 (15.90)	24.00 (6.93)
		20s	55.81 (15.03)	17.50 (4.60)
		Total	55.52 (12.06)	20.75 (6.62)
	Total	10s	59.85 (16.03)	20.07 (6.72)
		20s	58.75 (15.37)	17.53 (7.23)
		Total	59.30 (15.58)	18.80 (7.04)
low	NG	10s	61.96 (10.59)	16.90 (5.61)
		20s	55.04 (13.95)	19.90 (3.81)
		Total	58.50 (12.57)	18.40 (4.91)
	VGP1	10s	50.99 (16.12)	18.20 (4.37)
		20s	46.11 (11.33)	19.70 (4.57)
		Total	48.55 (13.79)	18.95 (4.42)
	VGP2	10s	50.06 (18.33)	19.00 (5.70)
		20s	43.84 (18.96)	20.70 (3.56)
		Total	46.95 (18.43)	19.85 (4.70)
	Total	10s	54.34	18.03

			(15.82)	(5.15)
		20s	48.33 (15.36)	20.10 (3.89)
		Total	51.33 (15.75)	19.07 (4.64)
Total	NG	10s	64.81 (15.50)	15.60 (6.97)
		20s	61.27 (15.11)	16.83 (6.86)
		Total	63.04 (15.28)	16.22 (6.89)
	VGP1	10s	56.08 (14.35)	17.80 (5.70)
		20s	55.32 (15.48)	17.17 (6.32)
		Total	55.70 (14.80)	17.48 (5.97)
	VGP2	10s	54.90 (15.79)	21.13 (5.84)
		20s	55.71 (17.15)	18.43 (4.73)
		Total	55.30 (16.34)	19.78 (5.44)
	Total	10s	58.60 (15.69)	18.18 (6.54)
		20s	57.43 (15.99)	17.48 (6.01)
		Total	58.01 (15.81)	17.83 (6.27)

Note. Standard Deviations are in parenthesis

7.4.5. SA Subscales:

The SA measure of SART is made up of attentional supply, demand and understanding. One way ANOVAs were conducted on these subscales to examine the effect of group and TL (see Table 30).

7.4.5.1. Demand: There was a main effect of TL on attentional demand F(2, 162) = 5.75, p = .004, η_p^2 = .07. Comparisons show that attentional demand was higher in high TL (*M* = 13.82, *SD* = 2.99) than the low TL (*M* =11.78, *SD*= 4.03), p = 009. Additionally, participants in the moderate TL (*M* = 13.72, *SD* = 4.03) found the task more demanding than low TL, p = .014. No differences between high TL and moderate TL, p = 1.00.

Additionally, significant differences were found for time conditions F(2, 162) = 5.17, p = .024, η_p^2 = .03. Participants reported more demand in 10 seconds (*M* = 12.73, *SD* = 4.02) than 20 seconds (*M* = 12.48, *SD* = 4.02). Hence, these findings demonstrate that when participants were given 10 seconds to make a decision they felt greater attentional demand. No differences in group F(2, 162) = .09, p = .922, η_p^2 = .00 were shown. There was also an interaction between TL and condition F(2, 162) = 3.41, p = .035, η_p^2 = .02. This interaction suggests that demand was lowest in low TL with 20 seconds to make a decision. Comparatively, it was highest in moderate TL with 10 seconds to make a decision. Hence, participant's perceptions of demand vary with different amounts of time to make a decision.

7.4.5.2. Supply: A main effect of TL on attentional supply F(2, 162) = 4.23, p = .016, η_p^2 = .05. Supply was higher in the moderate TL condition (*M* = 19.65, *SD* = 4.45) than high (*M* = 17.98, *SD* =3.74), p = .047 as well as compared to low (*M* = 17.88, *SD* = 2.64), p =.032. No differences between high and low, p = 1.00. Hence, the moderate TL increased feelings of attentional supply. Furthermore, a main effect of time condition to make a decision was found F(2, 162) = 3.89, p = .050, η_p^2 = .02. Participants reported a higher supply in 10 seconds (*M* = 19.06, *SD* = 4.10) than 20 seconds (*M* = 17.96, *SD* = 3.34). These results show that feelings of supply increase when participants were given 10 seconds to make a decision compared to 20 seconds.

7.4.5.3. Understanding: There was also a main effect of TL on understanding $F(1, 162) = 4.66$, $p = .011$, $\eta_p^2 = .05$. Individuals reported a higher understanding of the task in moderate TL (*M* = 13.17, *SD* = 4.13) than both high (*M* = 11.43, *SD* = 3.82), *p* = .020 and low TL (*M* = 13.03, SD = 3.38), *p* = .037. No difference between moderate and low TL, *p* = 1.00. Hence, individuals felt as though they had a poorer understanding of the task in high TL and the most in moderate TL. Further, a main effect was found with group $F(2, 162) = 7.82$, $p = .001$, $\eta_p^2 = .08$. NGs reported a lower understanding (*M* = 11.30, *SD* = 3.67) than VGP2 (*M* = 13.80, *SD* = 3.87), *p* < .001.Additionally, there was an interaction between TL, group and time condition $F(4, 162) = 2.57$, $p = .040$, $\eta_p^2 = .06$. NGs in 10s time condition and the high TL had the lowest understanding of the task.

Table 30: Means and Standard Deviations for dimensions of SA as influenced by TL, time and group

TL	Group	Time Condition	Demand	Supply	Understanding
High	NG	10s	12.80 (3.77)	17.00 (4.94)	8.70 (4.50)
		20s	13.50 (3.63)	16.50 (3.03)	10.70 (2.58)
		Total	13.15 (3.62)	16.75 (4.00)	9.70 (3.71)
	VGP1	10s	14.70 (2.83)	18.30 (4.69)	12.30 (3.43)
		20s	14.10 (1.97)	17.30 (3.23)	10.40 (3.06)
		Total	14.40 (2.39)	17.80 (3.96)	11.35 (3.31)
	VGP2	10s	13.70 (3.02)	20.50 (2.51)	13.60 (3.50)
		20s	14.10 (2.85)	18.30 (2.83)	12.90 (4.12)
		Total	13.90 (2.86)	19.40 (2.84)	13.25 (3.74)

	Total	10s	13.73 (3.22)	18.60 (4.30)	11.53 (4.26)
		20s	13.90 (2.81)	17.37 (3.02)	11.33 (3.40)
		Total	13.82 (2.99)	17.98 (3.74)	11.43 (3.82)
Mod	NG	10s	14.60 (4.35)	20.00 (5.46)	11.40 (4.17)
		20s	14.30 (4.99)	19.10 (3.78)	12.10 (3.70)
		Total	14.45 (4.56)	19.55 (4.59)	11.75 (3.85)
	VGP1	10s	14.00 (3.71)	19.40 (5.25)	14.70 (4.19)
		20s	12.70 (4.24)	19.60 (3.57)	12.50 (2.76)
		Total	13.35 (3.94)	19.50 (4.37)	13.60 (3.63)
	VGP2	10s	13.60 (2.72)	20.70 (5.31)	16.80 (2.53)
		20s	13.10 (4.56)	19.10 (4.18)	11.50 (4.88)
		Total	13.35 (3.66)	19.90 (4.72)	14.15 (4.66)
	Total	10s	14.07 (3.55)	20.03 (5.18)	14.30 (4.24)
		20s	13.37 (4.50)	19.27 (3.72)	12.03 (3.76)
		Total	13.72 (4.03)	19.65 (4.49)	13.17 (4.13)
Low	NG	10s	14.20 (3.05)	18.10 (2.38)	13.00 (2.71)
		20s	10.20 (4.05)	18.00 (2.45)	11.90 (3.28)

		Total	12.20 (4.05)	18.05 (2.35)	12.45 (2.98)
	VGP1	10s	13.40 (4.77)	18.40 (2.12)	13.20 (2.57)
		20s	9.50 (4.09)	17.10 (2.64)	12.10 (2.81)
		Total	11.45 (4.76)	17.75 (2.43)	12.65 (2.68)
	VGP2	10s	12.60 (3.66)	19.10 (2.13)	12.90 (2.88)
		20s	10.80 (2.86)	16.60 (3.69)	15.10 (3.31)
		Total	11.70 (3.33)	17.85 (3.20)	14.00 (3.23)
	Total	10s	13.40 (3.81)	18.53 (2.18)	13.03 (2.63)
		20s	10.17 (3.62)	17.23 (2.93)	13.03 (3.38)
		Total	11.78 (4.03)	17.88 (2.64)	13.03 (3.00)
Total	NG	10s	13.87 (3.71)	18.37 (4.49)	11.03 (4.15)
		20s	12.67 (4.49)	17.87 (3.21)	11.57 (3.17)
		Total	13.27 (4.13)	18.12 (3.88)	11.30 (3.67)
	VGP1	10s	14.03 (3.76)	18.70 (4.13)	13.40 (3.49)
		20s	12.10 (3.98)	18.00 (3.27)	11.67 (2.93)
		Total	13.07 (3.96)	18.35 (3.71)	12.53 (3.31)
	VGP2	10s	13.30 (3.09)	20.10 (3.56)	14.43 (3.37)

	20s	12.67 (3.67)	18.00 (3.64)	13.17 (4.28)
	Total	12.98 (3.38)	19.05 (3.72)	13.80 (3.87)
Total	10s	13.73 (3.51)	19.06 (4.10)	12.96 (3.91)
	20s	12.48 (4.02)	17.96 (3.34)	12.13 (3.55)
	Total	13.11 (3.81)	18.51 (3.77)	12.54 (3.75)

Note. Standard Deviations are in parenthesis.

7.4.6. NASA TLX:

There are 6 Subscales of WL; mental demand, physical demand, temporal demand, performance, effort and frustration. To identify a specific type of WL, ANOVAs were conducted to examine these subscales by TL and group (see Table 30).

7.4.6.1. Mental Demand: A main effect of TL on feelings of mental demand F(2, 162) = 4.89, $p = .009$, $\eta_p^2 = .06$. Mental demand was highest in high TL ($M = 295.08$, $SD = 96.65$) compared to both the moderate ($M = 240.42$, $SD = 119.60$), $p = .035$ and low TL conditions ($M = 234.33$, $SD = 131.53$), $p = .015$.

7.4.6.2. Temporal Demand: Significant differences were found in TL on temporal demand F(2, 162) = 10.13, $p < .001$, $\eta_p^2 = .11$. Comparisons show that individuals reported the high TL to be more temporally demanding ($M = 267.58$, $SD = 114.28$) compared to low ($M = 168.25$, $SD = 124.08$), $p < .001$. Moderate TL ($M = 225.83$, $SD = 124.76$) was also found to be more temporally demanding than low $p = .031$. No differences between high and moderate TL. Furthermore, a close to significant main effect of group F(2, 162) = 2.81, $p = .063$.

7.4.6.3. Performance: A main effect of group on feelings of performance was found F(2, 162) = 7.00, p = .001, η_p^2 = .08. NG reported higher levels of performance (M = 173.75, SD = 103.15) in the task than VGP2 (M = 111.10, SD = 85.92), p = .001. No significant differences Physical Demand, Effort, Frustration, p > .05.

Table 31: Means and Standard Deviations for dimensions of WL as influenced by TL, time condition and group

TL	Group	Time	Mental Demand	Physical	Temporal Demand	Performance	Effort	Frustration
High	NG	10s	299.00 (99.41)	1.00 (3.16)	282.00 (135.11)	212.00 (100.34)	105.50 (74.07)	138.00 (117.24)
		20s	261.5 (122.38)	.50 (1.58)	335.00 (116.17)	154.00 (121.01)	145.50 (91.18)	85.00 (96.26)
		Total	280.25 (110.21)	.75 (2.45)	308.50 (125.62)	183.00 (112.21)	125.50 (83.41)	111.50 (107.89)
	VGP1	10s	301.00 (89.84)	5.50 (11.41)	234.00 (91.65)	115.00 (35.28)	143.00 (91.99)	46.00 (45.02)
		20s	276.50 (126.69)	1.00 (3.16)	267.00 (58.18)	152.50 (112.40)	135.00 (91.86)	112.00 (82.77)
		Total	288.75 (107.63)	3.25 (8.47)	250.50 (76.60)	133.75 (83.33)	139.00 (89.57)	79.00 (73.15)
	VGP2	10s	326.50 (52.87)	7.00 (14.94)	255.00 (138.66)	76.50 (50.68)	157.00 (99.59)	69.50 (67.02)
		20s	306.00 (80.86)	8.00 (18.29)	232.50 (121.57)	117.50 (58.94)	204.00 (127.56)	144.50 (138.77)
		Total	316.25 (67.31)	7.50 (16.26)	243.75 (127.44)	97.00 57.48)	180.50 (113.96)	107.00 (112.83)
	Total	10s	308.83 (81.25)	4.50 (10.93)	257.00 (120.99)	134.50 (87.58)	135.17 (88.86)	84.50 (88.68)
		20s	281.33 (109.60)	3.17 (10.95)	278.17 (108.17)	141.33 (99.19)	161.50 (105.84)	113.83 (107.6)
		Total	295.08 (96.65)	3.83 (10.87)	267.58 (114.28)	137.92 (92.83)	148.33 (97.79)	99.17 (98.90)
Mod	NG	10s	194.00 (148.47)	.00 (.00)	293.50 (139.26)	166.50 (105.94)	119.00 (79.19)	177.50 (164.57)
		20s	276.50 (115.01)	13.00 (34.33)	180.00 (124.92)	178.00 (126.06)	183.00 (108.35)	135.00 (140.83)

		Total	235.25 (136.01)	6.50 (24.55)	236.75 (141.31)	172.25 (113.48)	151.00 (98.03)	156.25 (150.66)
	VGP1	10s	276.50 (142.73)	11.00 (28.46)	206.00 (129.95)	129.50 (57.95)	150.50 (110.39)	139.00 (131.19)
		20s	221.00 (103.25)	.00 (.00)	245.00 (122.11)	132.50 (85.58)	122.00 (62.55)	134.00 (126.35)
		Total	248.75 (124.54)	5.50 (20.38)	225.50 (124.35)	131.00 (71.15)	136.25 (88.54)	136.50 (125.38)
	VGP2	10s	280.50 (67.39)	3.00 (7.89)	221.00 (105.80)	94.00 (59.53)	141.00 (109.94)	89.50 (133.97)
		20s	194.00 (114.18)	9.00 (28.46)	209.50 (124.31)	121.50 (114.02)	143.00 (70.29)	160.50 (154.57)
		Total	237.25 (101.47)	6.00 (20.56)	215.25 (112.50)	107.75 (89.64)	142.00 (89.81)	125.00 (145.41)
	Total	10s	250.33 (127.35)	4.67 (17.12)	240.17 (127.45)	130.00 (80.81)	136.83 (98.28)	135.33 (143.73)
		20s	230.50 (112.61)	7.33 (25.45)	211.50 (122.46)	144.00 (108.89)	149.33 (83.98)	143.17 (136.68)
		Total	240.42 (119.60)	6.00 (21.55)	225.83 (124.76)	137.00 (95.33)	143.08 (90.85)	139.25 (139.11)
Low	NG	10s	303.50 (114.55)	0.00 (0.00)	215.50 (154.69)	156.50 (82.02)	135.00 (67.37)	120.40 (87.77)
		20s	223.00 (127.50)	0.00 (0.00)	199.00 (120.71)	175.50 (93.44)	116.50 (93.45)	115.00 (121.36)
		Total	263.25 (124.99)	0.00 (0.00)	207.25 (135.31)	166.00 (86.12)	125.75 (79.86)	117.70 (103.12)
	VGP1	10s	244.50 (136.09)	3.00 (9.49)	131.50 (147.22)	135.50 (96.19)	155.50 (85.32)	111.00 (98.51)
		20s	230.50 (150.95)	6.50 (18.86)	141.00 (80.03)	156.00 (80.06)	72.00 (57.70)	88.00 (63.78)
		Total	237.50 (140.06)	4.75 (14.64)	136.25 (115.43)	145.75 (86.77)	113.75 (82.82)	99.50 (81.63)
	VGP2	10s	214.50 (145.34)	1.00 (3.16)	180.50 (127.80)	143.00 (100.20)	99.50 (77.51)	115.00 (93.57)
		20s	190.00 (115.71)	1.50 (4.74)	142.00 (105.94)	114.10 (113.17)	115.50 (81.39)	95.00 (107.13)

		Total	202.25 (128.48)	1.25 (3.93)	161.25 (115.94)	128.55 (105.08)	107.50 (77.79)	105.00 (98.44)
		10s	254.17 (133.38)	1.33 (5.71)	175.83 (143.00)	145.00 (90.29)	130.00 (78.01)	115.47 (90.20)
	Total	20s	214.50 (128.81)	2.67 (11.20)	160.67 (103.70)	148.53 (96.71)	101.33 (79.02)	99.33 (97.62)
		Total	234.33 (131.53)	2.00 (8.84)	168.25 (124.08)	146.77 (92.78)	115.67 (79.18)	107.40 (93.54)
Total		10s	265.50 (128.95)	0.33 (1.83)	263.67 (142.60)	178.33 (96.43)	119.83 (72.16)	145.30 (125.11)
	NG	20s	253.67 (119.68)	4.50 (20.10)	238.00 (135.95)	169.17 (110.93)	148.33 (98.49)	111.67 (118.48)
		Total	259.58 (123.49)	2.42 (14.31)	250.83 (138.73)	173.75 (103.15)	134.08 (86.80)	128.48 (121.99)
		10s	274.00 (123.00)	6.50 (18.20)	190.50 (128.49)	126.67 (66.15)	149.67 (93.24)	98.67 (102.72)
	VGP1	20s	242.67 (126.37)	2.50 (11.04)	217.67 (103.87)	147.00 (91.07)	109.67 (75.03)	111.33 (93.32)
		Total	258.33 (124.64)	4.50 (15.06)	204.08 (116.64)	136.83 (79.58)	129.67 (86.30)	105.00 (97.51)
		10s	273.83 (104.97)	3.67 (9.91)	218.83 (124.37)	104.50 (76.36)	132.50 (96.44)	91.33 (100.20)
	VGP2	20s	230.00 (114.98)	6.17 (19.33)	194.67 (119.96)	117.70 (95.38)	154.17 (100.27)	133.33 (133.26)
		Total	251.92 (111.37)	4.92 (15.28)	206.75 (121.76)	111.10 (85.92)	143.33 (98.15)	112.33 (118.79)
		10s	271.11 (118.13)	3.50 (12.14)	224.33 (134.04)	136.50 (85.58)	134.00 (87.82)	111.77 (111.30)
	Total	20s	242.11 (119.47)	4.39 (17.19)	216.78 (120.60)	144.62 (100.63)	137.39 (93.09)	118.78 (115.35)
		Total	256.61 (119.36)	3.94 (14.85)	220.56 (127.18)	140.56 (93.23)	135.69 (90.25)	115.27 (113.08)

Note. Standard Deviations are in parenthesis

7.4.7. Between-Subjects Confidence-Accuracy:

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was also conducted. No significant relationship was found between-subjects confidence and accuracy $r(179) = -.102$, $p = .174$.

7.5. Discussion

The analysis reported in the Chapter was interested in examining whether there were any significant differences between individuals who were given 20 seconds to make a decision compared to those who received 10 seconds to make a decision. It aimed to provide insight into how different time pressures influence decision accuracy, confidence and metacognition. The results mirror the findings of the previous two experiments in regards to DC, TL and W-S C-A, and demonstrate that DC impacts on both decision accuracy and confidence. However, in general, the time to make a decision was found not to influence decision accuracy, confidence or W-S C-A in this comparison analysis. However, the impact of the time conditions was shown to influence attentional demand and supply on the subscale of SA. The main findings of task manipulation are also discussed.

7.5.1. Accuracy, Confidence and W-S C-A:

It was hypothesised that changing the time given to participants to make a decision would impact on decision accuracy, confidence and W-S C-A. However, the null hypotheses were accepted. Reducing the time to make a decision did not interfere with these variables. This supports previous research of Yang et al. (2012) who similarly found no effect of time pressure on accuracy, confidence or metacognition. However, the analysis conducted in this Chapter on the subscales of WL found no differences in reported feelings of temporal demand between 10 seconds and 20 seconds. Hence, the findings from the current analysis could be explained by the unsuccessful manipulation of time pressure. Nevertheless, as mentioned previously, it is unknown to researchers the time required to increase feelings of time pressure. Hence, future work could consider investigating different time pressures on these variables.

Significant findings were, however, reported for the impact of DC on decision accuracy. The comparison analysis also provides further support to the findings from the previous experiments. In line with the previous experiments, individuals were more accurate when making decisions that were high in criticality. This finding supports work which has demonstrated that cognition and performance are not impaired by criticality (Callister et al., 1999). In addition, this analysis showed that TL influenced decision accuracy. The analysis demonstrated that individuals were more accurate in the low TL condition compared to the moderate TL condition. Thus, in conditions of low cognitive and temporal load individuals were better able to respond to the decision events. This supports findings that higher load impairs cognition (Cumming & Harris, 2001; Arnsten, 2009). In summary, participants performed better when decisions were decisions were higher in criticality and in conditions of low TL. These findings may, therefore, suggest that there may be optimal conditions to make accurate decisions.

In regards to decision confidence, and in keeping with previous findings, VGP2 were, again, more confident in their decisions. However, as this was not matched to the accuracy of the decision, this would suggest that gamers who play more than 7 hours a week may apply too much confidence to a decision. In addition, time to make a decision did not impact individuals metacognitive abilities as reported by W-S C-A.

7.5.2. WL and SA:

Crucially, by examining the differences in WL and SA in the time to make a decision, research can build upon the time needed to induce feelings of time pressure. Previous research has also found that time pressure can increase the experienced WL of the task and has been linked to increased feelings of stress (Keinan et al., 1987). However, the analysis conducted found no differences in global WL and SA. By examining the subscales of WL, it was found that feelings of temporal demand (task speed) did not differ between 10 seconds and 20 seconds. This would suggest that the differences between the times given to make a decision were not diverse enough.

Main findings were demonstrated in the SA subscales. The analysis found differences in individual's feelings of attentional demand and supply of the task. Hence, participants who were given 10 seconds to make a decision rated attentional demand and supply higher than individuals who were given 20 seconds. The demand subscale is made up of the likeness of the situation will change, an increased number of variables to attend to and the complexity of the situation. Hence, increasing the time pressure increased the demand on an individual's attentional resources but, importantly, this did not influence decision accuracy, confidence, W-S C-A or WL. Furthermore, there was also an interaction between demand, time and TL. Demand was at its lowest in low TL with 20 seconds to make a decision. Comparatively, demand was highest in moderate TL with 10 seconds to make a decision. This suggests that reduced time to make a decision interacts with the TL to influence feelings of demand on the task. To investigate the impact time to make a decision has on SA it would be beneficial to conduct research using more objective measures of SA. In addition, the comparative analysis conducted found that feelings of supply increased when participants were given 10 seconds to make a decision compared to 20 seconds. Attentional supply refers to an individual's readiness for activity, the amount of mental availability for new variables, concentration and amount of division of attention arousal. An interaction was also observed between TL, group and supply. NGs in 10 seconds time condition and the high TL had the lowest understanding of the task. Hence, time pressure interacts with subjective feelings of SA by increasing feelings of supply and demand. Time pressure was not shown to impact on individuals understanding of the situation on the SA subscale. The results from this comparison found no significant differences between the time given to make a decision in the variables of decision accuracy, confidence, W-S C-A and WL. However, the manipulation of time pressure did increase the feelings of demand and supply which would suggest that SA is sensitive to changes in the time given to make a decision.

7.6. Summary

The purpose of this Chapter was to explore the impact of time pressure on decision making in an air defence scenario by comparing the results from Experiment 1 and Experiment 2. The results from this comparison demonstrate that time to make a decision applied here (i.e., 10 seconds versus 20 seconds)does not have an impact on decision accuracy, confidence and W-S C-A. Nor does it impact on overall WL and SA scores. However, the subscales suggest that 10 seconds to make a decision created more attentional demand and supply. It might be that a more systematic investigation of the time allowed to make decisions would add pressure. Nevertheless, the results do demonstrate that such pressures can impact on individual perceptions in relation to some SA subscales, but that this was neither detrimental nor beneficial to decision making. Further research would need to be conducted to examine the impact of the time pressures required to make a decision on decision accuracy, confidence and W-S C-A.

CHAPTER EIGHT

8. The Impact of Audio Attendance in Decision Accuracy, Confidence and W-S C-A

8.1. Introduction

The previous experimental Chapter examined the impact that time to make a decision had on the variables. Chapter eight presents the results from the third experiment which introduced an audio element to the task. The presence of audio in an Ops room is inevitable. Noise arises from radars, alerts, and most communication is conducted via conversation. Therefore, the aims of this experiment were two-fold. Firstly, to increase the fidelity of the experimental work, and secondly, to increase the understanding of the range of factors that influence decision making. Specifically, this study aimed to investigate the impact of the presence of background audio which is attended to as well as the presence of audio that does not have to be attended to. This was to ensure that participants did not simply just tune out of the audio communication as well as presenting an additional task for those attending.

Research suggests that the presence of audio can negatively impact on cognitive processing such as memory (Banbury & Ficker, 2003). Furthermore, as metacognition involves individuals monitoring their performance, audio may play a role in how individuals monitor their cognition (Finley, 2014). Indeed, research has found that interruptions to metacognitive ability reduce decision confidence but not necessarily accuracy (Beaman et al., 2014). Additionally, the introduction of an audio may also increase the feeling of WL as individuals have to divide their attention during the task. However, little is known of the impact of audio on SA in an air defence decision making task. Consequently, the findings from this experiment may also provide implications for decision aids which use auditory input to assist the decision maker (Vachon et al., 2011).

In summary, this experiment was interested in the impact of the presence of audio that was attended and audio that was not attended to. In line with the previous experiments, DC, TL and group were also assessed, as well as individual differences (Personality, cognitive constructs and video game play), and WL and SA.

8.2. Hypotheses

In addition to the hypotheses set out in Experiment 1:

- **HP13:** Audio attendance will influence decision accuracy, confidence and W-S C-A.
- **HP14**: Audio Attendance will influence feelings of WL and reduce SA.

8.3. Participants

Ninety participants were recruited through opportunity sampling from the University of Liverpool. The participants consisted of 31 females and 59 males with a mean age of 24 years (*SD* = 9.26). Participants were divided into 3 groups. Ethical approval and statistical criteria remained the same for this experiment (see Section 4.3.2 and 5.3).

8.4. Results

To assess the differences in means a number of statistical analyses were performed on the data for accuracy, confidence and W-S C-A using Analysis of Variance (ANOVA). A manipulation check was carried out to assess the differences in TL (see analysis of WL and SA). Significant differences were found in WL, p = .008 hence the TL manipulation was successful. There were no significant differences in SA in the different TL conditions. An alpha level of .05 was used for all statistical tests unless stated otherwise.

8.4.1. Accuracy:

The appropriateness and accuracy of the decisions was decided upon by the SMEs. When designing the decision log and generating the decision options one of the decision options was voted as being the best decision given the current situation. Participants scored '1' for a*n accurate response* or '0' for *an incorrect response*. The maximum total was 30 and the maximum mean for each DC was 10. To examine the mean differences between DC, TL and groups in accuracy an ANOVA was conducted (see Table 32).

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 2(Attend, not attend) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data.

A main effect of DC was found $F(2,144) = 44.45$, $p < .001$, $\eta_p^2 = .38$. Bonferroni corrected post hoc tests showed participants were more accurate in high DC decisions ($M = 5.13$ $SD = 1.70$) than both medium DC decision events ($M = 5.06$, $SD = 1.60$), $p < .001$ and low DC ($M = 3.30$, $SD = 1.56$), $p < .001$. Additionally, participants were more accurate in medium DC decisions than low DC decisions, $p < .001$. Hence, the most accurate decisions were made in the highly critical decisions.

A main effect of group was also shown $F(2,72) = 3.65$, $p = .031$, $\eta_p^2 = .09$. NG were significantly more accurate ($M = 4.75$, $SD = 1.28$) than VGP1 ($M = 4.06$, $SD = .71$) $p = .041$. No differences between NG and VGP2 ($M = 4.64$, $SD = 1.17$), $p = 1.00$ or VGP2 and VGP, $p = .120$. Additionally, no main effect of TL $F(2, 72) = 1.87$, $p = .162$, $\eta_p^2 = .05$ or audio $F(2,72) = 1.37$, $p = .245$, $\eta_p^2 = .02$ and no interaction effects were observed. Additionally, individuals who do not play games tended to be more accurate than those who play less than 7 hours a week (VGP1). Furthermore, neither TL nor audio attendance had an impact on decision accuracy.

Table 32: Means and Standard Deviations for Accuracy as influenced by DC, TL, audio and group

TL	Group	Audio	Overall	High DC	Medium DC	Low DC
High	NG	Attend	14.00 (2.24)	5.40 (1.14)	5.20 (1.14)	3.40 (2.07)
		Non Attend	12.40 (3.91)	4.60 (2.19)	4.60 (1.14)	3.20 (1.30)
		Total	13.20 (3.12)	5.00 (1.70)	4.90 (1.49)	3.30 (1.64)
	VGP1	Attend	10.80 (3.1)	4.00 (1.58)	4.20 (2.68)	2.60 (1.14)
		Non Attend	11.40 (3.83)	4.80 (1.92)	3.80 (1.10)	3.00 (.71)
		Total	11.10 (3.41)	4.40 (1.71)	4.00 (1.94)	2.80 (.92)
	VGP2	Attend	13.40 (3.05)	5.20 (1.64)	5.40 (1.52)	2.80 (1.30)
		Non Attend	15.40 (3.44)	5.40 (.55)	6.40 (2.30)	3.60 (1.34)
		Total	14.40 (3.24)	5.30 (1.16)	5.90 (1.91)	3.20 (1.32)
	Total	Attend	12.73 (3.22)	4.87 (1.51)	4.93 (1.98)	2.93 (1.49)
		Non Attend	13.07 (3.60)	4.93 (1.62)	4.93 (1.87)	3.27 (1.10)
		Total	12.90 (3.56)	4.90 (1.53)	4.93 (1.89)	3.10 (1.30)
Moderate	NG	Attend	15.67 (2.73)	5.67 (1.63)	6.17 (.98)	3.83 (2.04)

		Non Attend	12.50 (3.11)	4.75 (1.26)	5.50 (1.00)	2.25 (.96)
		Total	14.40 (3.17)	5.30 (1.49)	5.90 (.99)	3.20 (1.81)
	VGP1	Attend	11.00 (4.18)	3.00 (1.87)	4.40 (.89)	3.60 (2.17)
		Non Attend	12.80 (1.30)	4.60 (.55)	5.40 (1.32)	2.80 (1.48)
		Total	11.90 (3.07)	3.80 (3.80)	4.90 (1.20)	3.20 (1.93)
	VGP2	Attend	13.00 (3.67)	5.60 (1.67)	4.20 (.84)	3.20 (2.17)
		Non Attend	13.20 (2.78)	5.00 (1.87)	4.80 (1.30)	3.40 (1.52)
		Total	13.10 (3.07)	5.30 (1.70)	4.50 (1.08)	3.30 (1.77)
	Total	Attend	13.38 (3.85)	4.81 (2.04)	5.00 (1.27)	3.56 (2.07)
		Non Attend	12.86 (2.28)	4.79 (1.25)	5.21 (1.19)	2.86 (1.35)
		Total	13.13 (3.17)	4.80 (1.69)	5.10 (1.21)	3.23 (1.78)
Low	NG	Attend	14.00 (3.08)	6.40 (.89)	4.60 (2.07)	3.00 (1.41)
		Non Attend	17.00 (2.74)	7.40 (1.52)	5.60 (1.52)	4.00 (2.00)
		Total	15.50 (3.37)	6.90 (1.29)	5.10 (1.73)	3.50 (1.72)
	VGP1	Attend	11.80 (3.11)	3.60 (1.52)	4.40 (1.32)	3.80 (1.30)
		Non Attend	15.00 (4.24)	5.60 (1.52)	5.60 (2.04)	3.80 (2.39)

		Total	13.40 (3.37)	4.60 (1.78)	5.00 1.74)	3.80 (2.39)
	VGP2	Attend	13.75 (2.87)	5.25 (1.50)	5.00 (2.00)	3.50 (.58)
		Non Attend	14.67 (3.20)	5.83 (1.72)	5.50 (1.38)	3.33 (1.97)
		Total	14.30 (2.95)	5.60 (1.58)	5.30 (1.57)	3.40 (1.51)
	Total	Attend	13.14 (2.98)	5.07 (1.78)	4.64 (1.69)	3.43 (1.16)
		Non Attend	15.50 (3.37)	6.25 (1.78)	5.56 (1.55)	3.69 (1.99)
		Total	14.40 (3.37)	5.70 (1.78)	5.13 (1.56)	3.57 (1.63)
Total		Attend	14.63 (2.66)	5.81 (1.28)	5.38 (1.67)	3.44 (1.79)
		Non Attend	14.07 (3.79)	5.64 (2.10)	5.21 (1.25)	3.21 (1.58)
		Total	14.37 (3.19)	5.73 (1.68)	5.30 (1.47)	3.33 (1.68)
	VGP1	Attend	11.20 (3.49)	3.53 (1.60)	4.33 (1.68)	3.33 (1.68)
		Non Attend	13.07 (3.17)	5.00 (1.41)	4.93 (1.67)	3.20 (1.61)
		Total	12.13 (3.41)	4.27 (1.66)	4.63 (1.67)	3.27 (1.62)
	VGP2	Attend	13.36 (3.00)	5.36 (1.50)	4.86 (1.46)	3.14 (1.46)
		Non Attend	14.44 (3.08)	5.44 (1.46)	5.56 (1.71)	3.44 (1.55)
		Total	13.93 (3.04)	5.40 (1.45)	5.23 (1.61)	3.30 (1.49)

Total	Attend	13.09 (3.32)	4.91 (1.74)	4.87 (1.63)	3.31 (1.62)
	Non Attend	13.87 (3.32)	5.36 (1.65)	5.24 (1.55)	3.29 (1.55)
	Total	13.48 (3.33)	5.13 (1.70)	5.06 (1.60)	3.30 (1.56)

Note. Standard deviations are in parenthesis.

8.4.2. Confidence:

Participants were asked to rate confidence in their decision '0' being *not confident at all* and '5' *being extremely confident*. The maximum confidence score in total was 150 and for each DC 50. A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 2(Attend, not attend) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data. Mauchly's test of sphericity was found to be significant so Greenhouse-Geisser was used (see Table 33).

A main effect of DC was found F(1.5,111.4) = 9.77, p < .001, η_p^2 = .12. Bonferroni corrected post hoc tests showed that participants were significantly more confident in high DC decisions (M = 38.54, SD = 7.50) than medium DC decisions (M = 36.63, SD = 6.65), p < .001. Additionally, individuals were more confident in Low DC (M = 38.12, SD = 6.32) than medium DC, p < .001. No significant differences between low and high DC, p > .05. Hence, participants were equally as confident in high and low DC decisions.

A main effect of group was also observed F(2, 72) = 4.63, p = .013, η_p^2 = .11. VGP2 were significantly more confident (M = 40.27, SD =.70) than NG (M = 35.88, SD = 1.33) in their decisions p = .014. No differences were found between NG and VGP1 or VGP1 and VGP2, p > .05.

No main effect of audio attendance F(1, 72) =.10, p = .749, η_p^2 = .00. However, a significant interaction was observed between group and audio attendance F(2, 72) = 9.28, p < .001, η_p^2 = .21. VGP1 were less confident when required to attend to the audio compared to VGP2 who were more confident when required to attend the audio. NGs were more confident in decisions when asked to attend the audio than not to attend. In sum, DC impacts on decision confidence and groups display different levels of confidence, with VGP showing higher levels of decision confidence than NGs (Figure 12). Interactions between audio and TL and audio and DC were found to be significant, p > .05.

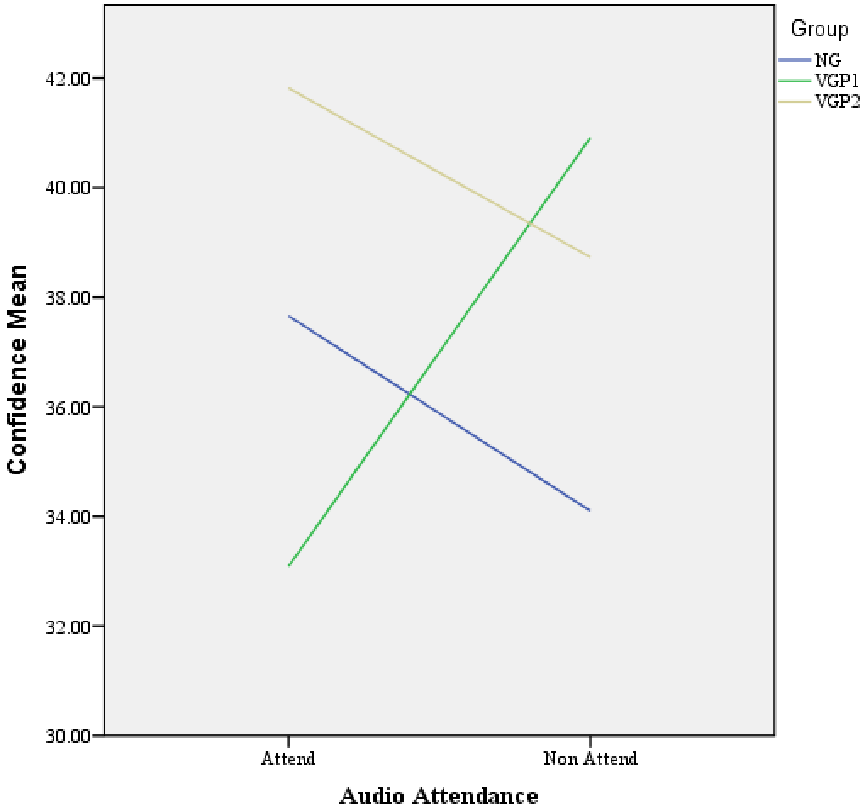


Figure 12. Graph displaying interaction between group and audio attendance in confidence

In addition, no main effect of TL on confidence was found F(1.72) = .37, p = .691, η_p^2 = .01. Hence, confidence did not vary between the TL conditions.

Table 33: Means and Standard Deviations for Confidence as influenced by DC, TL, group and audio

TL	Group	Audio	Overall	High DC	Medium DC	Low DC
High	NG	Attend	110.40 (15.18)	38.20 (4.66)	36.40 (5.03)	35.80 (7.66)
		Non Attend	106.00 (26.42)	35.40 (10.71)	34.00 (9.14)	36.60 (7.73)
		Total	108.20 (20.44)	36.80 (7.93)	35.20 (7.07)	36.20 (7.27)
	VGP1	Attend	97.40 (20.90)	33.20 (8.04)	31.80 (6.80)	32.40 (6.50)
		Non Attend	131.00 (6.89)	46.60 (2.07)	41.00 (3.00)	43.40 (2.61)
		Total	114.20 (23.00)	39.90 (8.98)	36.40 (6.93)	37.90 (7.45)
	VGP2	Attend	128.80 (15.40)	42.60 (5.18)	42.80 (4.71)	43.40 (6.19)
		Non Attend	117.20 (12.30)	40.20 (7.01)	39.60 (3.58)	37.40 (2.41)
		Total	123.00 (14.49)	41.40 (5.95)	41.20 (4.29)	40.40 (5.44)
	Total	Attend	112.20 (20.89)	38.00 (6.94)	37.00 (6.97)	37.20 (7.90)
		Non Attend	118.07 (19.19)	40.73 (8.41)	38.20 (6.32)	39.13 (5.53)

		Total	115.13 (19.19)	39.37 (7.70)	37.60 (6.56)	38.17 (6.77)
Moderate	NG	Attend	117.83 (25.02)	42.00 (8.17)	39.50 (9.20)	37.83 (7.49)
		Non Attend	99.50 (29.42)	31.00 (12.19)	31.50 (10.08)	37.00 (7.26)
		Total	110.50 (26.95)	37.60 (10.91)	36.30 (9.90)	37.50 (6.10)
	VGP1	Attend	97.20 (18.25)	32.40 (6.88)	32.00 (6.52)	32.80 (5.63)
		Non Attend	125.60 (12.56)	42.00 (4.73)	41.00 (3.54)	42.60 (5.41)
		Total	111.40 (21.10)	37.20 (7.53)	36.50 (6.85)	37.70 (7.33)
	VGP2	Attend	120.00 (7.94)	39.20 (5.54)	36.60 (4.67)	42.00 (4.95)
		Non Attend	108.20 (15.29)	35.60 (4.39)	35.60 (4.72)	37.00 (6.89)
		Total	114.10 (13.06)	37.40 (5.08)	36.10 (4.46)	39.50 (6.24)
	Total	Attend	112.06 (20.44)	38.13 (7.13)	36.25 (7.46)	37.56 (6.92)
		Non Attend	111.93 (21.10)	36.57 (2.92)	36.36 (7.07)	39.00 (6.10)
		Total	112.00 (20.44)	37.40 (5.31)	36.30 (7.15)	38.23 (6.69)

Low	NG	Attend	109.20 (15.22)	37.60 (7.13)	33.00 (5.61)	38.60 (3.78)
		Non Attend	101.40 (14.25)	35.00 (5.20)	31.40 (6.19)	35.00 (6.00)
		Total	105.30 (14.50)	36.30 (5.31)	32.20 (5.63)	36.80 (5.10)
	VGP1	Attend	103.20 (11.32)	35.00 (5.20)	33.40 (4.28)	34.80 (6.26)
		Non Attend	111.60 (18.51)	38.00 (7.55)	36.00 (6.81)	37.60 (4.62)
		Total	107.40 (15.13)	36.50 (6.31)	34.70 (5.45)	36.20 (5.39)
	VGP2	Attend	129.75 (23.96)	43.50 (11.09)	42.50 (6.61)	43.75 (7.14)
		Non Attend	123.17 (8.38)	44.00 (3.58)	40.17 (2.26)	39.00 (4.20)
		Total	125.80 (15.56)	43.80 (6.94)	41.10 (4.43)	40.90 (5.72)
	Total	Attend	112.93 (19.29)	38.36 (8.06)	35.86 (6.67)	38.71 (6.47)
		Non Attend	112.75 (15.94)	39.31 (6.11)	36.13 (6.23)	37.31 (4.91)
		Total	112.83 (17.30)	38.87 (6.98)	36.00 (6.33)	37.97 (5.64)

Total	NG	Attend	112.81 (18.66)	39.44 (6.77)	36.50 (7.15)	37.44 (6.29)
		Non Attend	102.50 (22.02)	34.00 (8.73)	32.36 (7.91)	36.14 (6.52)
		Total	108.00 (20.62)	36.90 (8.09)	34.57 (7.67)	36.83 (6.32)
	VGP1	Attend	99.27 (16.28)	33.53 (6.40)	32.40 (5.58)	33.33 (5.79)
		Non Attend	122.73 (15.11)	42.20 (6.10)	39.33 (5.04)	41.20 (4.84)
		Total	111.00 (19.51)	37.87 (7.56)	35.87 (6.30)	37.27 (6.60)
	VGP2	Attend	125.93 (15.11)	41.64 (7.06)	40.50 (5.72)	43.00 (5.63)
		Non Attend	116.63 (12.92)	40.19 (5.95)	38.56 (3.89)	37.88 (4.57)
		Total	120.97 (14.80)	40.87 (6.42)	39.47 (4.88)	40.27 (5.64)
	Total	Attend	112.38 (19.82)	38.16 (7.43)	36.38 (6.91)	37.80 (7.00)
		Non Attend	114.27 (18.53)	38.93 (7.64)	36.89 (6.45)	38.44 (5.62)
		Total	113.32 (19.10)	38.54 (7.50)	36.63 (6.65)	38.12 (6.32)

Note. Standard deviations are in parenthesis.

8.4.3. W-S C-A:

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Group: NG, VG1, VG2) X 2 (Attend, not attend) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data to assess individuals within-subjects confidence-accuracy (W-S C-A). No significant main effects or interactions were found in W-S C-A. Overall W-S C-A scores were very low and not negative (*M* = .09, *SD* = .18). The findings do not support the hypotheses TL, DC or audio would impact on W-S C-A. Due to the lack of significant findings, the results table reporting the means and standard deviation of W-S C-A as influenced by DC and TL are reported in Appendix 10c.

8.4.4. Percentage Confidence in Correct and Incorrect Responses:

To consider the variation in the data and examine the low correlations displayed for W-S C-A, percentage confidence in correct and incorrect responses was calculated. W-S C-A demonstrates the relationship between confidence and accuracy; however, a high correlation suggests both being highly confident in correct decisions as well as low confidence in incorrect decisions. Similarly, a negative correlation would suggest that individuals are highly confident in incorrect responses or not confident in correct responses. Thus, by examining the percentage confidence in incorrect or correct responses the direction of the confidence (over/under confidence) can be displayed (see Table 34).

To do this the number of correct responses was recorded and the confidence in those decisions calculated to produce percentage confidence in correct responses. The same method was applied to incorrect responses. Interestingly, all data suggests a high degree of confidence in decisions (*M* =75.55, *SD* = 12.73). Descriptive statistics show that means in correct and incorrect to be similar, individuals were slightly more confident in correct that in correct (*M* = 76.60, *SD* = 12.90), (*M* = 74.64, *SD* = 13.47).

Percentage Accuracy in correct responses. To examine the differences in TL and group a two way ANOVA was conducted on percentage confidence in correct responses. No main effect of TL on percentage confidence in correct decisions was observed $F(2, 72) = .36$, $p = .701$, $\eta_p^2 = .01$. A main effect was found on group $F(2,81) = 4.16$, $p = .019$, $\eta_p^2 = .10$. VGP2 (*M* = 81.51, *SD* = 9.05) were significantly more confident in correct responses than NG (*M* = 72.94, *SD* = 14.99), $p = .022$. No differences between NG and VGP1 (*M* = 72.63, *SD* = 16.37), $p = 1.00$ nor VGP1 and VGP2, $p = .109$. No main effect of audio $F(1, 72) = .07$, $p = .789$, $\eta_p^2 = .00$.

A significant interaction between group and audio $F(2, 72) = 8.10$, $p = .001$, $\eta_p^2 = .18$. Both NGs and VGP2 show higher confidence when attending audio. In comparison VGP1 show higher confidence when not attending No interaction other interactions were found to be significant $p < .05$.

Percentage Accuracy in incorrect responses. Similarly, an ANOVA was conducted with percentage confidence in incorrect responses. *Levene's test was found to be significant.* Again, no main effect percentage incorrect was found in TL $F(2, 81) = .51$, $p = .605$, $\eta_p^2 = .01$. However, a main effect of group was found to be significant $F(2, 81) = 4.80$, $p = .011$, $\eta_p^2 = .12$. Differences were observed between NGs (*M* = 70.31, *SD* =14.25) and VGP2 (*M* = 80.02, *SD* = 11.10), $p = .010$. No differences between NGs and VGP1 (*M* = 73.51, *SD* = 13.60) $p = .965$, or VGP1 and VGP2, $p = .133$. No main effect of audio $F(1, 72) = .01$, $p = .925$, $\eta_p^2 = .00$. Furthermore, an interaction between group and audio was observed $F(2,72) = 8.91$, $p < .001$, $\eta_p^2 = .20$. In a similar manner to incorrect VGP1 were least confident when attending than attending. In comparison both NGs and VGP2 showed a reduction in confidence when not attending the audio No other interactions were observed, $p > .05$.

Table 34: Means and Standard Deviations for % confidence (correct and incorrect) according to TL, group and audio

TL	Group	Audio	% Correct	% Incorrect
High	NG	Attend	76.23 (10.12)	71.73 (11.60)
		Non Attend	71.16 (17.17)	69.27 (18.54)
		Total	73.69 (13.55)	70.50 (14.64)
	VGP1	Attend	65.90 (13.04)	64.88 (14.66)
		Non Attend	87.46 (6.18)	88.76 (7.70)
		Total	76.68 (14.89)	76.82 (16.74)
	VGP2	Attend	83.94 (11.69)	86.73 (9.84)
		Non Attend	79.00 (6.47)	77.38 (10.56)
		Total	81.47 (9.28)	82.06 (10.81)
	Total	Attend	75.35 (13.25)	74.45 (14.72)
		Non Attend	79.20 (12.43)	78.47 (14.59)

		Total	77.28 (12.77)	76.46 (14.59)
Mod	NG	Attend	82.47 (18.58)	76.34 (16.29)
		Non Attend	65.65 (20.84)	66.97 (18.76)
		Total	75.74 (20.30)	72.59 (16.98)
	VGP1	Attend	65.41 (14.10)	64.86 (12.10)
		Non Attend	81.93 (11.68)	83.77 (8.28)
		Total	73.67 (14.99)	74.31 (13.61)
	VGP2	Attend	80.17 (7.42)	79.29 (7.62)
		Non Attend	73.99 (9.74)	71.36 (10.98)
		Total	77.08 (8.79)	75.32 (9.81)
	Total	Attend	76.42 (15.57)	73.68 (13.50)
		Non Attend	74.44 (14.72)	74.54 (13.92)
		Total	75.50 (14.95)	74.32 (13.47)

Low	NG	Attend	72.61 (11.18)	73.32 (9.23)
		Non Attend	69.54 (7.53)	64.23 (12.65)
		Total	71.08 (9.13)	68.77 (11.98)
	VGP1	Attend	71.02 (6.65)	67.18 (8.68)
		Non Attend	77.66 (12.53)	71.59 (10.58)
		Total	74.34 (10.08)	69.39 (9.41)
	VGP2	Attend	87.24 (13.77)	86.07 (17.78)
		Non Attend	84.71 (4.69)	79.29 (7.64)
		Total	85.72 (8.78)	82.00 (12.25)
	Total	Attend	76.22 (12.19)	74.77 (13.59)
		Non Attend	77.77 (10.30)	72.17 (11.85)
		Total	77.04 (11.05)	73.39 (12.53)

Total	NG	Attend	77.44 (13.93)	73.96 (12.29)
		Non Attend	69.00 (14.62)	66.81 (15.79)
		Total	73.50 (14.65)	70.62 (14.25)
		Attend	67.44 (11.18)	65.64 (11.23)
	VGP1	Non Attend	82.35 (10.58)	81.37 (11.15)
		Total	74.90 (13.11)	73.51 (13.60)
		Attend	83.53 (10.55)	83.89 (11.55)
	VGP2	Non Attend	79.58 (8.053)	76.22 (9.67)
		Total	81.42 (9.35)	79.79 (11.10)
		Attend	76.00 (13.51)	74.27 (13.63)
	Total	Non Attend	77.21 (12.38)	75.01 (13.45)
		Total	76.61 (12.90)	74.64 (13.47)

Note. Standard deviations are in parenthesis.

8.4.5. WL and SA:

A one-way ANOVA was conducted to assess the relationship between SA and TL, audio and group. There was no significant main effect of TL or group on SA. A significant interaction was observed between group and audio attendance in reported SA F(2, 72) = 4.01, *p* = .022, η_p^2 = .10. This interaction implies that VGP2s who were instructed to attend to the audio reported higher levels of SA than VGP1s who were also instructed to attend the audio. NGs reported similar between attending and not attending. VGP1 reported higher SA when attending than not attending. No other main effects or interactions were found to be significant (see Table 35).

A one-way ANOVA was conducted to assess the relationship between WL and TL, audio, and group. Significant differences were found in TL F(2, 72) = 5.24, *p* = .008, η_p^2 = .13. This finding suggests that the TL manipulation was successful. WL was significantly higher in the high TL (*M* = 60.64, *SD* = 16.71) than low TL (*M* = 52.63, *SD* = 14.73), *p* = .028 and moderate was rated higher in TL (*M* = 63.92, *SD* = 12.80) than low TL, *p* = .002.

Furthermore, significant differences were also found in between the groups F(2,72) = 3.41, *p* = .039, η_p^2 = .09. Comparisons show that NG (*M* = 62.36, *SD* = 14.26) rated WL to be higher than VGP2 (*M* = 53.70, *SD* = 15.47), *p* = .018. VGP1 (*M* = 61.12, *SD* = 15.60) also rated WL as higher than VGP2, *p* = .040. This finding suggests that individuals that do not play video games found the task more demanding than those that play video games for more than 7 hours a week.

Significant differences were also found between those who attended the audio and those that did not F(2,72) = 9.18, *p* = .003, η_p^2 = .11. Comparisons show that attending the audio resulted in reporting higher levels of WL (*M* = 63.60, *SD* = 14.35) than those who did not attend (*M* = 54.63, *SD* =15.21), *p* = .003.The finding suggesting that audio attendance increases individual's feelings of WL. No significant interaction effects were observed.

Table 35: Means and Standard Deviations for WL and SA as influenced by TL, group and audio

TL	Group	Audio	Overall WL	Overall SA
High				
	NG	Attend	70.76 (10.19)	18.20 (5.96)
		Non Attend	65.50 (15.22)	18.40 (5.18)
		Total	68.13 (12.52)	18.30 (5.27)
	VGP1	Attend	76.10 (12.83)	18.40 (3.51)
		Non Attend	52.10 (13.07)	16.60 (6.19)
		Total	64.10 (17.58)	17.50 (4.84)
	VGP2	Attend	51.36 (17.47)	19.80 (3.21)
		Non Attend	48.04 (13.73)	19.60 (3.98)
		Total	49.70 (14.92)	19.70 (3.43)
	Total	Attend	66.07 (16.88)	18.80 (4.16)
		Non	55.21 (15.11)	18.20 (4.97)

		Attend		
		Total	60.64 (16.69)	18.50 (4.52)
Moderate	NG	Attend	67.53 (16.71)	18.67 (5.92)
		Non Attend	65.88 (17.00)	20.00 (4.08)
		Total	66.87 (66.87)	19.20 (5.05)
	VGP1	Attend	69.58 (9.19)	13.80 (5.72)
		Non Attend	58.80 (11.30)	23.20 (6.14)
		Total	64.19 (11.25)	18.50 (7.47)
	VGP2	Attend	62.64 (9.31)	21.60 (2.30)
		Non Attend	59.10 (14.00)	14.60 (5.13)
		Total	60.87 (11.37)	18.10 (5.26)
	Total	Attend	66.64 (12.14)	18.06 (5.67)
		Non Attend	60.93 (13.02)	19.21 (6.17)

		Total	63.98 (12.80)	18.60 (5.83)
Low	NG	Attend	52.38 (7.27)	20.20 (4.38)
		Non Attend	52.16 (10.64)	19.20 (7.40)
		Total	52.27 (8.59)	19.70 (5.76)
	VGP1	Attend	65.30 (8.64)	11.80 (5.89)
		Non Attend	44.84 (17.66)	16.20 (6.57)
		Total	55.07 (17.97)	14.00 (6.33)
	VGP2	Attend	55.80 (21.18)	21.25 (3.30)
		Non Attend	45.28 (16.71)	17.50 (5.01)
		Total	49.49 (12.90)	19.00 (4.62)
	Total	Attend	57.97 (12.94)	17.50 (6.22)
		Non Attend	47.29 (12.76)	17.63 (6.00)

		Total	52.28 (14.76)	17.57 (6.00)
Total		Attend	63.81 (14.15)	19.00 (5.20)
		Non Attend	60.84 (14.77)	19.14 (5.42)
		Total	62.42 (14.27)	19.07 (5.21)
	VGP1	Attend	70.33 (10.66)	14.67 (5.56)
		Non Attend	51.91 (12.46)	18.67 (6.72)
		Total	61.12 (15.61)	16.67 (6.39)
	VGP2	Attend	56.66 (15.47)	20.86 (2.85)
		Non Attend	50.46 (15.27)	17.25 (4.88)
		Total	53.35 (15.42)	18.93 (4.39)
	Total	Attend	63.76 (14.35)	18.13 (5.31)
		Non Attend	54.18 (15.21)	18.31 (5.64)

	Total	58.97 (15.47)	18.22 (5.45)

Note. Standard deviations are in parenthesis.

8.4.6. SA Subscales:

To examine the three dimensions of SA as measured by SART (Demand, Supply, and Understanding) one-way ANOVAs were carried out across each dimension and TL condition (see Table 36).

8.4.6.1 Demand: No main effect of TL, $F(2, 72) = .70$, $p = .501$, $\eta_p^2 = .02$, group $F(2, 72) = .42$, $p = .661$, $\eta_p^2 = .01$ or audio $F(2, 72) = 1.57$, $p = .214$, $\eta_p^2 = .02$. A significant interaction between TL and group was found $F(4, 72) = 3.48$, $p = .012$, $\eta_p^2 = .16$. The three different groups show similar demand in the high TL. NG found the moderate TL condition considerably more demanding compared to the low TL. VGP2 reported similar levels of demand across conditions. VGP1 reported lower demand in moderate TL which increased demand in the low TL condition (Figure 13). No other interactions were significant. Hence, gamers show a similar trend compared to the NGs.

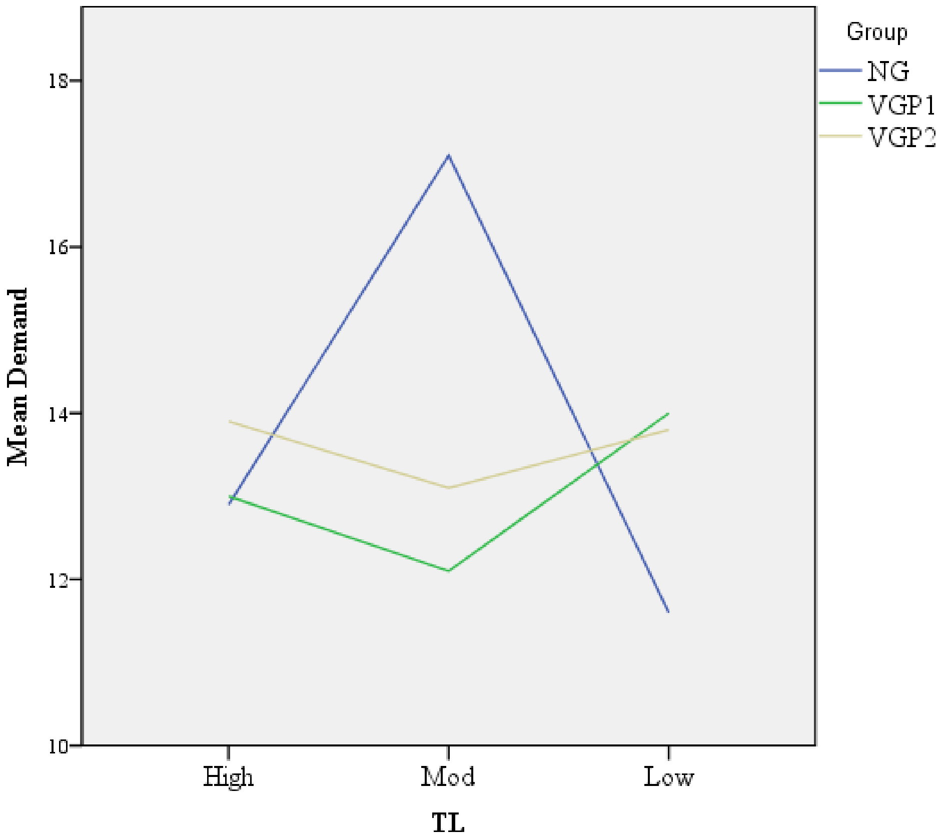


Figure 13. Graph displaying interaction between group and TL in Demand subscale

8.4.6.2. Supply: In the supply subscale, no main effects were found for TL, $F(2,72) = 1.39$, $p = .256$, $\eta_p^2 = .04$, group $F(2,72) = 1.37$, $p = .260$, $\eta_p^2 = .04$ or audio $F(2, 72) = .002$, $p = .967$, $\eta_p^2 = .00$. However, there was a significant interaction between TL and group $F(4, 72) = 3.01$, $p = .023$, $\eta_p^2 = .14$. NG found the moderate TL required more attentional supply than the high TL. Although similar across high and low, VGP1 rated moderate and low requires less supply, and VGP2 rated similar levels of supply for high and low and lower supply levels in the moderate TL condition (Figure 14).

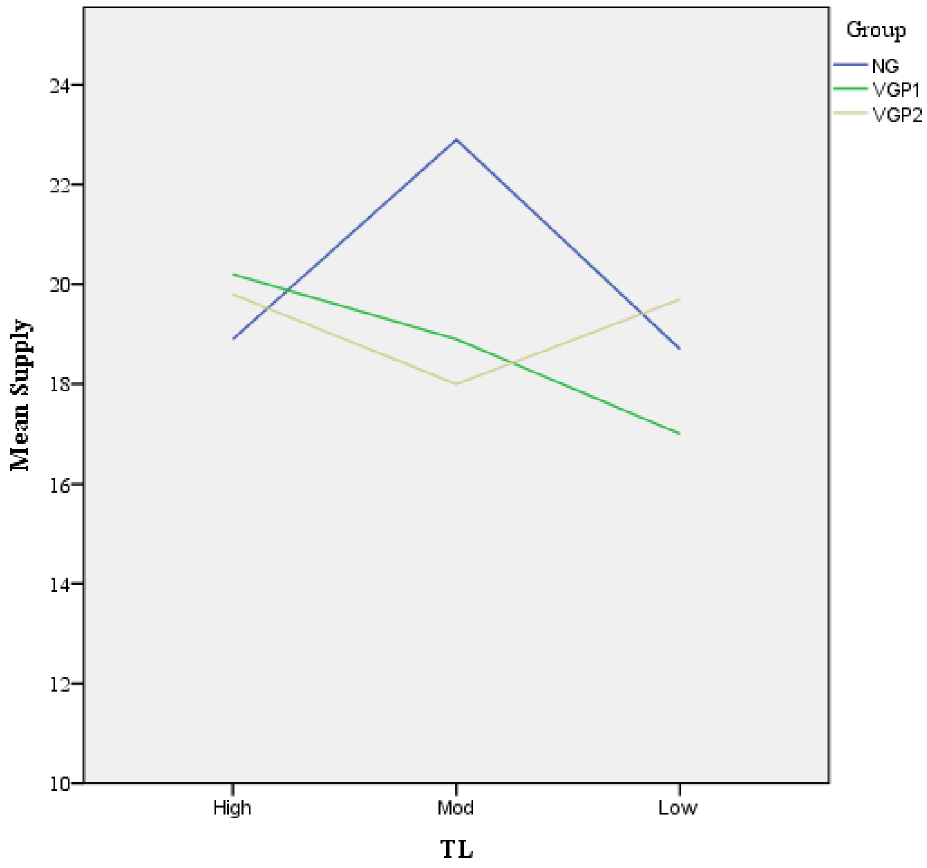


Figure 14. Graph displaying interaction between group and TL in Supply subscale

8.4.6.3. Understanding: With regards to subscale of understanding a main effect of group was found $F(2, 72) = 3.78$, $p = .028$, $\eta_p^2 = .10$. As such, VGP2 reported higher level of understanding (*M* = 13.74, *SD* = 2.76) than VGP1 (*M* = 11.40, *SD* = 3.65), $p = .023$. No differences between NG and VGP1, $p = .369$ and NG and VGP2, $p = .732$. No other effects were found to be significant. Finding suggests that those who played more hours of video games believed they had more understanding of the task.

Table 36: Means and Standard Deviations for dimensions of SA as influenced by TL, group and audio

TL	Group	Audio	Demand	Supply	Understanding
High					
	NG	Attend	14.40 (4.93)	19.40 (3.91)	13.20 (4.66)
		Non Attend	11.40 (3.58)	18.40 (3.78)	11.40 (1.82)
		Total	12.90 (4.36)	18.90 (3.67)	12.30 (3.47)
	VGP1	Attend	13.60 (4.56)	20.00 (3.16)	12.00 (4.69)
		Non Attend	12.40 (2.70)	20.40 (3.64)	11.00 (3.16)
		Total	13.00 (3.59)	20.20 (3.65)	11.50 (3.81)
	VGP2	Attend	13.60 (5.68)	18.40 (4.51)	15.00 (1.14)
		Non Attend	14.20 (3.42)	21.20 (2.95)	14.60 (2.97)
		Total	13.90 (4.43)	19.80 (3.88)	14.80 (2.20)
	Total	Attend	13.87 (4.72)	19.27 (3.67)	13.40 (3.83)
		Non Attend	12.67 (3.24)	20.00 (3.70)	12.33 (3.02)

		Total	13.27 (4.03)	19.63 (3.67)	12.87 (3.43)
Moderate	NG	Attend	17.67 (3.20)	22.50 (3.67)	13.83 (4.83)
		Non Attend	16.25 (1.26)	23.50 (4.20)	12.75 (2.22)
		Total	17.10 (2.60)	22.90 (3.70)	13.40 (3.86)
	VGP1	Attend	13.00 (2.83)	17.80 (2.78)	9.00 (2.45)
		Non Attend	11.20 (3.56)	20.00 (4.18)	14.40 (4.88)
		Total	12.10 (3.18)	18.90 (3.54)	11.70 (4.62)
	VGP2	Attend	12.80 (3.77)	20.20 (1.30)	14.20 (2.28)
		Non Attend	13.40 (1.82)	15.80 (2.78)	12.20 (3.70)
		Total	13.10 (2.81)	18.00 (3.09)	13.20 (3.08)
	Total	Attend	14.69 (3.88)	20.31 (3.32)	12.44 (4.07)
		Non Attend	13.43 (3.11)	19.50 (4.70)	13.14 (3.40)
		Total	14.10 (3.54)	19.93 (3.97)	12.77 (3.85)

Low					
	NG	Attend	12.20 (2.39)	19.00 (1.23)	13.40 (2.51)
		Non Attend	11.00 (2.55)	18.40 (5.32)	11.80 (2.78)
		Total	11.60 (2.41)	18.70 (3.65)	12.60 (2.63)
	VGP1	Attend	15.20 (3.11)	16.60 (4.72)	10.40 (2.07)
		Non Attend	12.80 (1.30)	17.40 (2.30)	11.60 (3.13)
		Total	14.00 (2.58)	17.00 (3.53)	11.00 (2.58)
	VGP2	Attend	12.75 (2.22)	20.25 (2.75)	13.75 (2.36)
		Non Attend	14.50 (3.62)	19.33 (3.08)	12.67 (3.26)
		Total	13.80 (3.12)	19.70 (2.83)	13.10 (2.85)
	Total	Attend	13.43 (2.79)	18.50 (3.39)	12.43 (2.65)
		Non Attend	12.88 (2.64)	18.44 (3.58)	12.06 (2.91)
		Total	13.13 (2.85)	18.47 (3.47)	12.23 (2.75)

Total	Attend	14.94 (4.12)	20.44 (3.42)	13.50 (3.91)
	Non Attend	12.64 (3.46)	19.86 (4.79)	11.93 (2.20)
	Total	13.87 (3.94)	20.17 (4.05)	12.77 (3.27)
VGP1	Attend	13.93 (3.43)	18.13 (3.68)	10.47 (3.29)
	Non Attend	12.13 (2.59)	19.27 (3.75)	12.33 (3.85)
	Total	13.03 (3.14)	18.70 (3.70)	11.40 (3.64)
VGP2	Attend	13.07 (3.95)	19.57 (3.06)	14.36 (1.95)
	Non Attend	14.06 (2.93)	18.81 (3.54)	13.13 (3.26)
	Total	13.60 (3.42)	19.17 (3.29)	13.70 (2.75)
Total	Attend	14.02 (3.85)	19.40 (3.47)	12.76 (3.54)
	Non Attend	12.98 (3.05)	19.29 (3.96)	12.49 (3.17)
	Total	13.50 (3.49)	19.34 (3.70)	12.62 (3.35)

Note. Standard deviations are in parenthesis.

8.4.7. NASA TLX:

As an exploratory analysis, the 6 dimensions of the NASA TLX were also examined to determine whether differences existed across the conditions. One way ANOVAs were conducted with TL across the different dimensions of workload (see Table 37).

8.4.7.1. Mental Demand: For mental demand there were no main effects of group F(2,72) = .47, p = .625, η_p^2 = .01, audio F(1,72) =3.52, p = .065, η_p^2 = .05 or TL, F(2, 72) = 1.98, p =.146, η_p^2 = .05. No interactions were observed for group and audio F(2, 72) = 1.74, p = .182, η_p^2 = .05, or group, audio and TL, F(4,72) = 2.38, p = .251, η_p^2 = .07. However, there was a significant interaction between group and TL F(4,72) = 3.44, p = .017, η_p^2 = .15. VGP2 rate the demand to be higher in the moderate TL in comparison NGs and VGP1 rate the demand to be highest in the high TL. NGs report the demand to be lowest in the low TL compared to VGP1 who display higher levels of demand in the low TL (Figure 15). Results indicate that gaming experience influences the perception of perceived mental demand of the task.

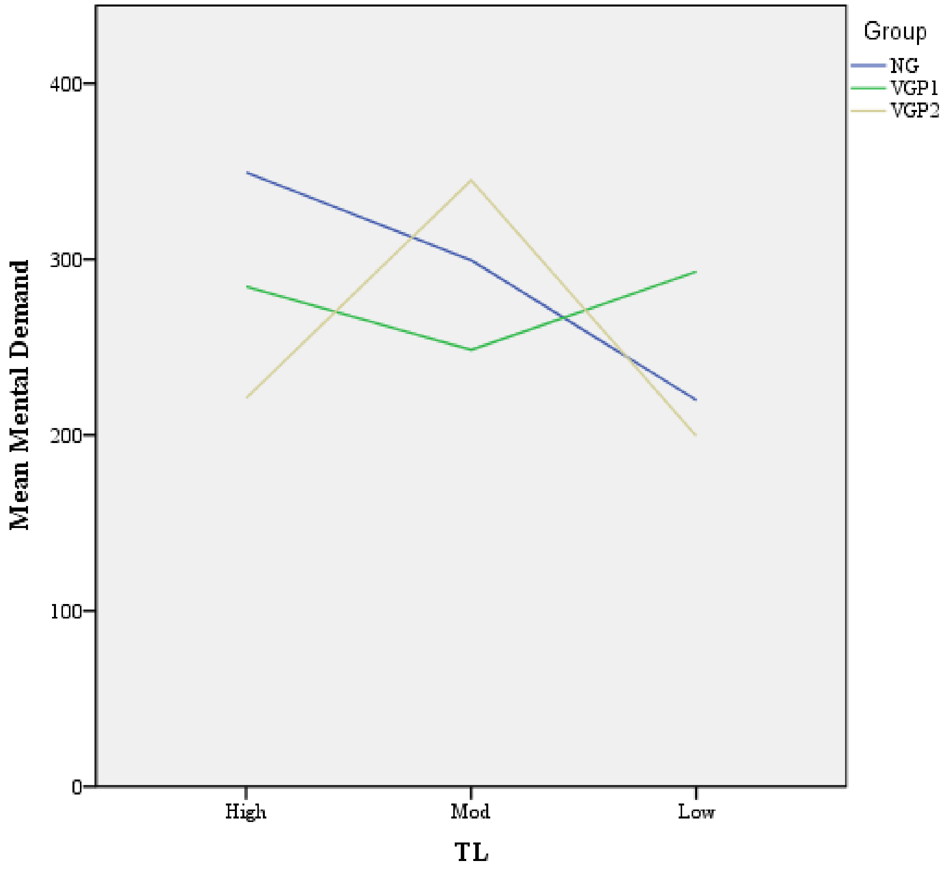


Figure 15. Graph displaying interaction bewtween group and TL in Mental Demand Subscale.

8.4.7.2. Effort: *Levene's test was found to be significant*. A main effect of TL was found on effort F(2, 72) =3.63, p = .032, η_p^2 = .09. Differences were found between low TL (*M* = 107.43, *SD* = 81.31) and moderate TL (*M* = 163.17, *SD* = 95.34), p = .049 which suggests that individuals felt that more effort was required in the moderate TL condition than the low TL. No significant differences between high and low, p = .096 or high and moderate TL, p = 1.00.

8.4.7.3. Temporal Demand: For temporal demand no main effects were observed in group F(2,72) = 1.43, p = .246, η_p^2 = .04, audio F(1,72) =.1.05, p =.308, η_p^2 = .01 or TL F(2, 72) = 2.57, p = .084. η_p^2 = .07. No interactions were observed between group and audio F(2, 72) = .70, p = .499, η_p^2 = .02, group and TL F(4,72) = 1.61, p = .181, η_p^2 = .08, group, audio and TL F(4,72) = .410, p = .801, η_p^2 = .02.

No main effects or interactions found in performance, frustration or physical demand p > .05.

Table 37: Means and Standard Deviations for dimensions of WL as influenced by to TL, group and audio

TL	Group	Audio	Mental Demand	Physical demand	Temporal Demand	Performance	Frustration	Effort
High								
	NG	Attend	363.00 (76.04)	0.00 (.00)	252.00 (133.12)	153.00 (133.77)	146.00 (77.94)	148.00 (66.38)
		Non Attend	336.00 (68.78)	0.00 (.00)	196.00 (563.38)	169.00 (120.85)	114.00 (148.64)	168.00 (58.90)
		Total	349.50 (69.82)	0.00 (.00)	224.00 (102.63)	161.00 (120.48)	130.00 (113.16)	158.00 (60.10)
	VGP1	Attend	349.00 (95.81)	0.00 (.00)	366.00 (101.08)	122.00 (69.88)	92.00 (60.58)	203.00 (118.20)
		Non Attend	220.00 (130.42)	5.00 (11.18)	269.00 (68.78)	110.00 (60.42)	41.00 (30.29)	137.00 (107.56)

		Total	284.50 (127.53)	2.50 (7.91)	317.50 (96.21)	116.00 (61.91)	66.50 (52.55)	170.00 (112.08)
	VGP2	Attend	219.00 (169.02)	0.00 (.00)	235.00 (58.31)	99.00 (82.80)	112.00 (151.10)	106.00 (76.76)
		Non Attend	233.00 (150.69)	43.00 (93.38)	148.00 (104.08)	104.00 (58.99)	24.00 (42.78)	179.00 (93.43)
		Total	221.00 (150.98)	21.50 (66.25)	191.50 (91.80)	101.50 (67.83)	68.00 (114.51)	142.50 (89.33)
	Total	Attend	310.33 (130.16)	0.00 (.00)	284.33 (112.15)	124.67 (94.82)	116.67 (99.21)	152.33 (92.87)
		Non Attend	259.67 (125.79)	16.00 (54.06)	204.33 (90.81)	127.67 (84.45)	59.67 (93.44)	161.33 (84.44)
		Total	285.00 (128.38)	8.00 (38.43)	244.33 (108.20)	126.17 (88.24)	88.17 (99.03)	156.83 (87.33)
Moderate								
	NG	Attend	364.17 (69.53)	5.83 (10.21)	240.00 (76.16)	130.00 (89.44)	90.83 (103.85)	182.50 (126.32)
		Non Attend	202.50 (59.09)	37.50 (47.87)	322.50 (119.55)	123.75 (30.92)	120.00 (162.53)	182.50 (143.32)
		Total	299.50 (104.01)	18.50 (33.00)	273.00 (99.00)	127.50 (69.09)	102.50 (122.57)	182.50 (125.37)
	VGP1	Attend	252.00 (127.01)	10.00 (14.14)	289.00 (62.29)	184.00 (149.10)	183.00 (137.91)	116.00 (93.96)
		Non Attend	245.00 (140.71)	1.00 (2.24)	279.00 (163.65)	123.00 (60.37)	25.00 (29.37)	209.00 (56.72)
		Total	248.50 (126.43)	5.50 (10.66)	284.00 (116.85)	153.50 (111.95)	104.00 (125.58)	162.50 (88.07)
	VGP2	Attend	300.00 (78.74)	0.00 (.00)	297.00 (129.11)	138.00 (89.26)	45.00 (28.50)	160.00 (93.34)

		Non Attend	390.00 (76.24)	0.00 (.00)	215.00 (99.62)	108.00 (85.78)	45.00 (27.16)	129.00 44.22)
		Total	345.00 (87.11)	0.00 (.00)	256.00 (117.00)	123.00 (84.04)	45.00 (26.25)	144.50 (70.77)
	Total	Attend	309.06 (99.47)	5.31 (10.24)	273.13 (90.15)	149.38 (106.36)	105.31 (110.42)	154.69 (104.00)
		Non Attend	284.64 (125.05)	11.07 (28.84)	268.57 (128.97)	117.86 (60.53)	59.29 (90.85)	172.86 (87.22)
		Total	297.67 (110.82)	8.00 (20.87)	271.00 (108.00)	134.67 (88.03)	83.83 (102.72)	163.17 (95.34)
Low								
	NG	Attend	231.00 (118.40)	0.00 (.00)	247.00 (151.15)	138.00 (77.59)	81.00 (62.69)	89.00 (50.67)
		Non Attend	209.00 (127.79)	11.00 (24.60)	251.00 (45.61)	146.00 (43.79)	85.00 (109.20)	81.00 (76.52)
		Total	220.00 (116.71)	5.50 (17.39)	249.00 (105.27)	142.00 (59.45)	83.00 (83.97)	85.00 (61.33)
	VGP1	Attend	347.00 (82.81)	7.00 (15.62)	163.00 (119.46)	103.00 (70.05)	162.00 (69.88)	198.00 (110.09)
		Non Attend	239.00 (17.66)	0.00 (.00)	197.00 (105.69)	79.00 (40.68)	56.00 (65.52)	102.00 (58.91)
		Total	293.00 (113.09)	3.50 (11.07)	180.00 (107.83)	91.00 (55.47)	109.00 (84.85)	150.00 (97.41)
	VGP2	Attend	223.75 (200.45)	12.50 (25.00)	198.75 (155.48)	201.25 (82.20)	100.00 (161.92)	126.25 (81.38)
		Non Attend	183.33 (134.11)	0.00 (.00)	191.67 (155.23)	163.33 (122.42)	93.17 (164.33)	48.33 (38.17)

		Total	255.17 (134.03)	5.00 (15.81)	194.50 (146.51)	178.50 (104.70)	95.90 (154.12)	79.50 (68.09)
	Total	Attend	270.36 (110.79)	6.07 (15.71)	203.21 (135.51)	143.57 (81.18)	115.36 (100.62)	138.57 (91.64)
		Non Attend	208.75 (121.72)	3.44 (13.75)	211.88 (111.00)	131.56 (85.69)	79.00 (116.60)	75.31 (59.26)
		Total	237.50 (118.93)	4.67 (14.50)	207.83 (120.92)	137.17 (82.40)	95.97 (109.34)	104.83 (81.31)
Total								
		Attend	322.19 (104.51)	2.19 (6.57)	245.94 (113.04)	139.69 (95.61)	105.00 (84.24)	142.50 (93.72)
		Non Attend	252.50 (107.07)	14.64 (31.04)	251.79 (88.93)	147.86 (75.21)	105.36 (129.68)	141.07 (94.55)
		Total	289.67 (109.72)	8.00 (22.23)	248.67 (100.81)	143.50 (85.33)	105.17 (145.67)	141.83 (94.55)
	VGP1	Attend	316.00 (106.71)	5.67 (12.08)	272.67 (124.93)	136.33 (102.13)	145.67 (97.46)	172.33 (108.83)
		Non Attend	234.67 (121.74)	2.00 (6.49)	248.33 (116.72)	104.00 (54.06)	40.67 (43.67)	149.33 (85.69)
		Total	275.33 (119.84)	3.83 (9.71)	260.50 (119.43)	120.17 (81.95)	93.17 (91.42)	160.83 (96.54)
	VGP2	Attend	249.29 (120.70)	3.57 (13.36)	246.79 (116.13)	142.14 (88.98)	84.64 (119.54)	131.07 (81.20)
		Non Attend	260.31 (148.46)	13.44 (52.43)	185.31 (119.77)	127.50 (93.32)	56.50 (103.03)	114.38 (80.89)
		Total	255.17 (134.03)	8.83 (39.08)	214.00 (120.16)	134.33 (90.06)	69.63 (110.01)	122.17 (80.07)
	Total	Attend	297.44 (112.82)	3.78 (10.77)	255.11 (116.04)	139.33 (93.70)	112.22 (101.52)	148.89 (94.67)

Non Attend	249.33 (125.45)	10.00 (35.61)	277.00 (112.00)	126.00 (76.73)	66.42 (99.70)	134.33 (87.75)
Total	273.39 (121.07)	6.89 (26.35)	241.06 (114.27)	132.67 (85.42)	89.32 (102.66)	141.61 (91.06)

Note. Standard deviations are in parenthesis.

8.4.8. Audio Attendance:

At the end of the task, participants were asked four simple questions based on the audio that was played to them (see Appendix 8d). A one way ANOVA revealed no significant differences between groups that were asked to attend and those were not instructed to attend $F(1,89) = 3.09$, $p = .082$. Although close to significance, this finding suggests the Audio Attendance manipulation was not successful. However, the descriptive statistics display that those in the attend condition did answer more questions correctly ($M = 2.04$, $SD = 1.07$) compared to non-attend ($M = 1.64$, $SD = 1.09$).

8.4.9. Between-Subjects Confidence Accuracy:

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was also conducted, no significant relationship was found between-subjects confidence and accuracy, $r(88) = -.01$, $p = .922$

8.4.10. Relationships between WL, SA and Accuracy, Confidence and W-S C-A:

A number of Pearson's correlations were carried out to establish whether a relationship existed between WL, SA, accuracy, confidence and W-S C-A. A correction was applied for multiple correlations. An alpha level of .005 was used.

Significant negative relationship between WL and confidence was found $r(88) = -.34$, $p = .001$. The higher the WL reported the less confident individuals were in their decisions. A close to significant positive relationship between SA and confidence was found $r(88) = .27$, $p = .011$. Similarly, the higher SA reported the more confident individuals were in their decisions. These findings demonstrate that decision confidence plays a vital role in both WL and SA. No significant differences were found in WL or SA in between accuracy. W-S C-A was not related to WL or SA.

However, close to significant relationships were found between overall accuracy and agreeableness $r(88) = .27$, $p = .010$. Individuals that scored higher on the agreeableness scale were more accurate in their decisions. No other psychometrics related to accuracy. A close to significant relationship was also found between confidence and neuroticism $r(88) = -.260$, $p = .013$. No other psychometrics measures were significant $p > .005$. W-S C-A scores were not related to any psychometric scores $p > .005$.

To assess the relationship between WL and SA, a series of Pearson's correlations were calculated. No significant relationship was found between SA and WL, $r(88) = -.17$, $p = .\ 108$.

To investigate whether there were individual differences in participants experiences of WL and SA correlations were conducted on each measure of the NEO-PI-R. WL was found to be positively related to agreeableness $r(88) = .26$, $p = .014$ and conscientiousness $r(88) = .27$, $p = .008$. No others were found to be significant and psychometric scores were not related to SA $p > .005$.

8.4.11. Personality Constructs:

To assess and compare each group on the psychometric measures, one-way ANOVAs were performed on the data (see Table 38). One way ANOVAs were conducted for the measures. The only significant main effect was found on group on scores of conscientiousness $F(2, 89) = 4.59$, $p = .013$, $\eta_p^2 = .\ 10$. Comparisons show that NG were more conscientiousness ($M = 32.67$, $SD = 6.21$) than VGP2 ($M = 27.30$, $SD = 7.01$), $p = .003$. No other comparisons were found to be significant.

Table 38: Means and Standard Deviations for dimensions of psychometric scores as influenced by group

Measure	NG	VGP1	VGP2	Total
Tolerance to Ambiguity	47.03 (7.21)	48.60 (6.25)	49.27 (6.23)	48.30 (6.57)
Decisiveness	18.87 (5.66)	21.50 (5.22)	21.70 (4.94)	20.69 (5.38)
Ambiguity Tolerance B	33.17 (7.62)	35.53 (5.64)	37.07 (6.10)	35.26 (6.63)
Decision Style	53.20 (10.67)	57.10 (9.49)	58.83 (9.53)	56.38 (10.08)
Neuroticism	20.93 (9.05)	23.97 (9.67)	26.03 (9.08)	23.64 (9.40)
Extraversion	31.43 (6.62)	29.67 (5.18)	29.00 (7.33)	30.03 (6.45)
Openness To experience	34.47 (6.36)	32.83 (5.90)	32.53 (6.57)	33.28 (6.27)
Agreeableness	34.43 (7.50)	31.63 (7.40)	30.53 (6.32)	32.20 (7.20)
Conscientiousness	32.67 (6.21)	29.60 (7.37)	27.30 (7.01)	29.86 (7.15)

Note. Standard deviations are in parenthesis.

8.5. Discussion

The aims of this experiment were two-fold. To increase the fidelity of the Experimental Work, and to increase the understanding of the range of factors that influence decision making. As such, this experiment reported in this Chapter introduced an auditory element to the decision making task. Specifically, the experiment aimed to investigate whether the introduction of an audio element had an impact on decision accuracy, confidence and metacognition. All other manipulations of TL, DC were kept the constant in line with Experiment 1, as well as the individual difference measures and WL and SA measurements. The results from this experiment demonstrated that audio did not influence decision accuracy, confidence or W-S C-A. However, audio which was attended to increased subjective feelings of global WL. Further, this experiment provided support to the findings in both experiments one and two. Consequently, the results from this study found that DC impacted on both decision accuracy and confidence. VGP2s were also again found to be more confident but not more accurate in their decisions. Additionally, in this experiment, individual differences in the personality traits of neuroticism and agreeableness were found to be related to performance, and conscientiousness to performance.

8.5.1. Accuracy:

In accordance to the previous findings in the first experiment (see Chapter 5) and second experiment (see Chapter 6) individuals in this task were significantly more accurate in the high DC and medium DC decisions than the low DC. This finding provides further support that DC influences the accuracy of decisions. Hence, the criticality of decisions is a crucial external factor to influence air defence decision making. Additionally, again, in line with Experiment 1, TL did not impact on the accuracy of the decision as there were no significant differences across high, moderate or low TL conditions.

However, unlike the previous experiments, a difference in decision accuracy was found between groups. NGs were found to be marginally more accurate than VGP1s. VGP1 had the lowest performance in this task. Hence, it could be argued from this finding that the accuracy of decisions, when audio is present is impacted by the amount of time spent playing video games; although no differences were found between VGP2 and NGs. As such, future research needed to look at these differences in the length of time playing games and the interactions on decision making in more detail. For instance, one study found that the length of time playing games influenced risky decisions (Bailey et al., 2013).

With regards to the audio manipulation, no differences in accuracy were found in individuals who attended and those who were instructed not to attend. This would suggest that having audio present did not influence the individual's performance in this task. To assess whether the manipulation was successful a set of questions were asked to the participants at the end of the task. Those in the condition to attend were aware of these questions at the start of the task, those in the not attend condition were not aware that there would be questions presented to them after the task. The audio manipulation did not find any significant differences in the number of correct decisions between those who attended the audio and those who did not. This finding would suggest that the audio manipulation was not successful. An explanation for this finding could be that individuals chose not to attend to the audio. This could be explained by the fact that the audio may not have been processed as it was not related to the task. Indeed, a higher perceptual load has been found to use more attentional capacity, reducing processing for task-irrelevant information (Lavie, 1995). In support of this, further analysis of SA subscales of attentional supply, which includes feelings of divided attention, showed no differences between the task conditions in feelings in individuals attending and not attending. This would suggest that that the individuals may not have engaged with the audio manipulation.

8.5.2. Confidence:

In this experiment, in line with the first two experiments, individuals were again more confident in high DC and low DC events than medium DC. This finding supports the previous experiments which demonstrated that individuals were the least confident in medium DC events. Decision confidence remained the same throughout the different TL conditions. This supports the finding from the previous experiments that decision confidence was not influenced by TL. With regards to the groups, VGP2s were also found to be more confident in their decisions than VGP1 and NGs. Additionally, the results from the current experiment also found a significant interaction between group and audio. VGP1 were less confident when required to attend to the audio compared to VGP2 and NGs who were more confident when required to attend the audio. VGP2 showed very high levels of confidence when asked to attend to the audio. This finding may imply that based on previous experience, as many games have some element of audio involved, this familiarity increase confidence in the decision (Wheatcroft et al., 2017).

8.5.3. W-S C-A:

In accordance with Experiment 1, neither DC, TL, nor did group had an influence on metacognitive ability. In addition, the presence of audio was also found not to influence individuals W-S C-A score. Furthermore, metacognitive ability remained low with individuals displaying high levels of overconfidence in this task as reporting in the percentage confidence analysis. The level of overconfidence also provides an explanation for the low W-S C-A findings in the current experiment.

8.5.4. WL and SA:

WL and SA were again assessed as performance measures. No differences were found in global SA in TL, group or audio conditions. However, an interaction was found between group and audio in reports of SA. The results found that VGP2s that were instructed to attend to the audio reported higher levels of SA than VGP1 that were also instructed to attend the audio. It could be that attending the audio increased VGP2s belief in SA. As such, it has been demonstrated that gaming and presence of audio are beneficial to feelings of SA (Chiappe et al., 2013).

Significant differences in WL were found between the TL conditions which would indicate that the manipulation of TL was successful in the experiment. The findings show that WL was reported to be higher in high TL compared to the low TL. Furthermore, differences between the groups in subjective feelings of WL were also found. Supporting the findings of Experiment 2, NGs reported higher WL than both VGP1 and VGP2. Research has shown that video gaming can increase attentional resources (Boot et al., 2008) and attentional visual field (Hubert-Wallander et al., 2011) hence, with the audio present; VGPs may be better suited to the demands of the task than NGs. It is therefore argued that these findings of increased SA and reduced WL in VGPs provide some evidence that gamers may be a better-suited population than students in understanding decision making in critical environments. As such helping to provide insight into decision making process and bridge a gap between novice and experts knowledge (see discussion in Chapter 12).

Additionally, a key finding in this experiment was that audio attendance increased feelings of WL. As previously mentioned, audio has been found to increase task load as it applies a higher perceptual load on the individual (Lavie, 1995). This has crucial implications as research has examined ways in which to reduce the WL of the operators by designing systems which provide audio information. However, in this experiment, although the instruction to attend to audio increased the perceived WL of the decision makers it was also found to be unrelated to decision accuracy, confidence and W-S C-A. Thus, although WL increased with the additional instruction of the requirement to attend to the audio, this experience did not have an adverse impact on the decisions made.

In regards to the impact of WL on decision accuracy and confidence, in agreement with the first experiment, the findings also demonstrate that high decision confidence was related to lower WL and higher decision confidence was related to higher reported SA. As mentioned, the manipulation of audio was not successful as there were no significant differences in the responses to questions at the end of the task. However, the descriptive statistics show that those attending the audio did answer more questions correctly. Hence, a limitation of this experiment could arise from the manipulation of audio attendance. Future research should examine different methods to manipulate how audio is attended to experimentally.

8.5.5. Personality and Cognitive Constructs:

The experiment also considered individual differences in personality and cognitive constructs. The experiment supported the findings of the first and second experiment, which found significant differences between groups. VGP2s were again also found to be less conscientious. Similarly to the previous experiments, this experiment found further evidence that personality constructs might be related to decision confidence and decision accuracy in this task. Although, these results did not meet the new significance level, there was evidence for trends in the data. As such, the findings from this experiment found a positive relationship between accuracy and agreeableness. Individuals who scored high on agreeableness had higher levels of decision accuracy, and hence made fewer errors, in the task. In addition, individuals who score more highly on neuroticism tend to be less emotionally stable, have a tendency to be anxious and display higher levels of worry. Neuroticism was found to negatively impact on performance in this task with lower accuracy scores relating to individuals who scored higher in neuroticism. An explanation for this finding could relate to research which has shown that cognition is impaired in highly neurotic individuals (Bryne, Silasi-Mansay & Worthy, 2015; Matthews, Deary & Whiteman, 2003) and could therefore provide a marker for recruitment into critical domains. Additionally, the personality trait of neuroticism was negatively related to overall confidence and therefore high feelings of distress and worry may reduce confidence in decisions. These findings have implications for individuals who are best suited to deal with the demands certain job roles, such as those in air defence.

Unlike the previous experiments, this experiment also demonstrated individual differences in the experience of WL. WL was found to be positively related to agreeableness and conscientiousness. Such a finding implies that certain characteristics may impact on feelings of WL (Chiorri et al., 2015).

8.6. Summary

The results from this experiment replicated the consistent finding that DC influence decision accuracy and confidence as previously demonstrated in the previously reported experiments. Additionally, the introduction of an audio element to the task was found not to impact of decision accuracy, confidence and W-S C-A. However, the manipulation was also found to be unsuccessful. Nevertheless, this experiment did find WL to be influenced by the audio manipulation which has implications on how decision support may be provided to operators.

SECTION B: Application

CHAPTER NINE

9. Investigating the Impact of Metacognitive Feedback Training (MFT) on Decision Accuracy, Confidence and W-S C-A

9.1. Introduction

As previously stated in Chapter one, decision confidence has implications on a course of action (Kepecs & Mainen, 2012). Therefore, it could be assumed that should confidence attributed to the decision be incorrect this could lead to an incorrect course action. Hence, an operator's ability to correctly identify their accuracy or inaccuracy in a decision is important. Thus far, the experimental Chapters in the thesis demonstrated relatively low W-S C-A scores - which was generally attributed to overconfidence in responses. That is, individuals seem to be applying higher levels of confidence to incorrect responses. These findings support a well-established finding in decision making literature (Wheatcroft, Wagstaff & Kebbell, 2004; Tyersky & Kahneman, 1974) mentioned in Chapter one. However, as stated this has implications on future actions (Desender et al., 2018). In order to achieve this, individuals must already have an awareness of their metacognitive ability. Poor metacognitive ability in an individual has been linked to individual's lack of awareness of their understanding of a task, as well as limiting their abilities to maximise performance. These are important aspects of any training. To recall, metacognition includes metacognitive awareness, which involves the monitoring and control of one's own performance and regulation and awareness of their ability (Flavell, 1979). Indeed, the ability to monitor one's own performance has been found to be successfully linked to improved learning in an educational setting (Tanner, 2012). Hence, there is research to suggest that metacognitive ability can be a trained skill (Jøsok et al., 2016; Kim, 2018).Thus, as well as examining the current W-S C-A relationship, consideration as to whether metacognitive skills could be trained in an air defence environment was investigated.

Recent research has begun to investigate the impact of training metacognition in an air defence domain. A study by Kim (2018) found that feedback has a positive training impact and improved individual's metacognitive judgements in some of the tasks. Furthermore, Morewedge et al. (2015) found that games that produce personalised feedback and practice had both short term and long term effects on reducing biases in decision making. Nevertheless, research is lacking in understanding the potential external and internal factors which may relate to individuals improvement in metacognitive monitoring and the use of other methods in this domain. For instance, research into personality has demonstrated a link between some personality types and better training success (Flin, 2012). Subsequently, there may be individual differences with regards to the benefits of metacognitive feedback (Dunlosky & Metcalfe, 2008). Furthermore, it has also been argued that cognitive and situational factors can determine the effects of training in a range of difference biases (Poos, Van den Vosch & Janssen, 2017). However, little is known about this in regards to air defence decision making. As previously noted SA and WL have also been linked to effective decision making. Therefore, the improvement of metacognitive ability may also influence individuals SA and WL. Indeed, studies by Cohen et al. (1998) and Poos et al., (2017) found that training in critical skills improved SA. However little is known about the influence of feedback on WL.

Hence, in addition to understanding the factors that may be involved in the relationship between confidence and accuracy, the fourth experiment in the current thesis investigated whether metacognitive instruction and feedback provided to individuals before the task could improve metacognitive ability in a decision making task. In summary, this next experiment was interested in applying a technique of improving metacognitive feedback via a short PowerPoint and instructional technique. Consequently, a metacognitive feedback training (MFT) Tool was developed. Prior to the task, individuals in the feedback training condition were provided with information to increase understanding of the C-A relationship combined with an additional briefing provided during the practice phase (i.e., which included personalised feedback after a practice trial of the decision making task). The present study examined the impact of metacognitive instruction on metacognitive monitoring ability in an air defence decision making task.

9.2. Hypotheses

Based on the previous experiments the HP are as follows:

- **HP14:** DC high DC will produce higher accuracy scores (A) and low DC higher confidence (B) and no impact of W-S C-A (C).
- **HP15**: No differences in groups with accuracy (A). VGPs will be significantly more confident in their decisions (B), no differences in W-S C-A between groups (C).
- **HP16:** MFT will increase accuracy (A), reduce confidence (B) and individuals will have a higher W-S C-A relationship in MFT condition (C).
- **HP17**: There will be differences in WL and SA in groups that received MFT and non-feedback groups.

9.3. Participants

Thirty participants were recruited through opportunity sampling from The University of Liverpool to take part in the MFT condition. These participants were compared with 30 participants randomly selected in SPSS from Experiment 1. In total 60 participants were used in this analysis. Mean age = 26.26 (*SD* = 7.49). Additionally, in this experiment, the sample consisted of two groups which were made up of VGPs and NGs. VGPs collapsed the previous groups of VGP1 and VGP2 and an even number of VGP1 and VGP2 formed the VGP group used in this study. Ethical approval and statistical criteria remained the same for this experiment (see Section 4.3.2 and 5.3).

9.4. Materials

Metacognitive Feedback Training (MFT) was provided to participants in MFT condition. A PowerPoint briefing was shown to the participants (see Appendix 8e). The PowerPoint briefing included information about the relationship between C-A in more detail to help individuals apply correct levels of confidence to perceived levels of accuracy as well as providing an example from another context (see full details in Chapter 4).

9.5. Procedure

The procedure followed the same course as in previous experiments apart from the addition or not of MFT. Participants who were randomly allocated to the MFT condition were instructed on the C-A relationship prior to the task (see Appendix 8e). In addition, after the practice trial, participants in the MFT condition went through their responses from the practice with the researcher. The researcher let them know if they had made a correct or incorrect decision and whether this was meaningfully indicated by the confidence score provided. The moderate TL caused the most uncertainty from the previous experiments and maintained the lowest W-S C-A. As such, all participants were assigned to the moderate TL condition and only the moderate TL condition from Experiment 1 was compared.

9.6. Results

To assess the differences in means a number of statistical analyses were performed on the data for accuracy, confidence and W-S C-A using Analysis of Variance (ANOVA). An alpha level of .05 was used for all statistical tests unless stated otherwise.

9.6.1. Accuracy:

A 2 (MFT, Yes, No) X 2 (Group: NG, VGP) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data (see Table 39).

In accordance with the hypotheses, a significant main effect of DC was found F(2, 112) = 17.64, $p < .001$, $\eta_p^2 = .24$. Individuals were significantly more accurate in high DC (*M* = 4.77, *SD* = 1.61) compared to low DC (*M* = 3.35, *SD* = 1.81), $p < .001$ and significantly more accurate in medium (*M* = 5.08, *SD* = 1.85) than the low DC, $p < .001$. No differences between high DC and medium DC, $p = .739$. Overall, these results suggest that individuals are more accurate in medium DC decisions.

There was no main effect of MFT on accuracy of decisions F(1,56) = .00, $p = 1.00$, $\eta_p^2 = .00$. Hence, receiving feedback did not improve performance on the task. Additionally, group did not impact on the accuracy of the decisions F(1.56) = .88, $p = .353$, $\eta_p^2 = .02$. Furthermore, no interactions with DC and accuracy or interaction between feedback and gamers were observed. Subsequently, in regards to accuracy in the decisions, only DC had an impact, with both high and moderate DC improving the accuracy of decisions.

Table 39: Means and Standard Deviations for Accuracy as influenced by MFT, group and DC

MFT	Group	Overall accuracy	High DC	Medium DC	Low DC
Yes	VGP	13.07 (3.51)	4.53 (1.73)	5.53 (2.39)	3.00 (1.46)
	NG	13.40 (3.50)	4.40 (1.55)	5.07 (1.44)	3.87 (2.03)
	Total	13.23 (3.45)	4.47 (1.61)	5.30 (1.95)	3.43 (1.79)
No	VGP	12.60 (3.20)	4.93 (1.28)	4.60 (1.64)	3.00 (1.89)
	NG	13.73 (3.17)	5.20 (1.86)	5.13 (1.89)	3.53 (1.85)
	Total	13.17 (3.18)	5.07 (1.57)	4.87 (1.76)	3.27 (1.86)

	VGP	12.83 (3.31)	4.73 (1.51)	5.07 (2.07)	3.00 (1.66)
Total	NG	13.57 (3.29)	4.80 (1.73)	5.10 (1.65)	3.70 (1.91)
	Total	13.20 (3.29)	4.77 (1.61)	5.08 (1.85)	3.35 (1.81)

Note. Standard deviations are in parenthesis.

9.6.2. Confidence:

Participants were asked to rate confidence in their decision ‘0’ being *not confident at all* and ‘5’ *being extremely confident*. The maximum confidence score in total was 150 and for each DC 50. A 2 (MFT, Yes, No) X 2 (Group: NG, VGP) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data (see Table 40). Mauchly’s test of sphericity was found to be significant - as such Greenhouse-Geisser *df* is reported.

Again, supporting the hypothesis, a main effect of DC was observed $F(2, 87.06) = 5.75$, $p = .008$, $\eta_p^2 = .09$. Comparisons reveal that individuals were more confident in the high DC decisions ($M = 39.02$, $SD = 6.45$) than medium DC ($M = 37.08$, $SD = 6.25$), $p < .001$ and marginally more confident in low DC ($M = 38.47$, $SD = 7.18$) than medium DC, $p = .061$. Again the finding supports the hypothesis that high DC increases confidence in decisions. Furthermore, there was no impact on MFT $F(1,56) = 2.31$, $p = .134$, $\eta_p^2 = .04$ and no significant interactions were observed, $p > .05$.

There was however a significant impact of group $F(1, 56) = 8.71$, $p = .005$, $\eta_p^2 = .14$. VGPs were significantly more confident in their decisions ($M = 40.36$, $SD = 5.47$) than NG ($M = 36.01$, $SD = 6.99$). No interaction between MFT and group was observed, $p > .05$. Subsequently, these findings demonstrate that decision confidence was impacted by DC and game play. As a result, higher levels of decision confidence were found in either high or low DC with medium DC impairing decision confidence.

Additionally, individuals who play video games applied higher confidence to their decisions.

Table 40: Means and Standard Deviations for Confidence as influenced by MFT, group and DC

MFT	Group	Overall Confidence	High DC	Medium DC	Low DC
Yes	VGP	124.67 (10.53)	42.20 (4.55)	40.00 (3.80)	42.47 (3.78)
	NG	111.67 (17.04)	37.47 (6.13)	36.07 (6.77)	37.67 (5.88)
	Total	118.17 (15.41)	39.83 (5.83)	38.03 (5.75)	40.07 (5.43)
No	VGP	117.80 (17.50)	41.13 (6.26)	38.33 (5.47)	38.07 (7.60)
	NG	104.53 (21.31)	35.27 (6.68)	33.93 (7.21)	35.67 (9.17)
	Total	111.17 (20.22)	38.20 (7.02)	36.13 (6.67)	36.87 (8.36)
Total	VGP	121.23 (14.62)	41.67 (5.40)	39.17 (4.70)	40.27 (6.31)
	NG	108.10 (19.21)	36.37 (6.40)	35.00 (6.95)	36.67 (7.63)
	Total	114.67 (18.17)	39.02 (6.45)	37.08 (6.25)	38.47 (7.18)

Note. Standard deviations are in parenthesis.

9.6.3. W-S C-A:

To assess W-S C-A a 2 (MFT: Yes, No) X 2(Group: NG, VGP) X 3(Decision Criticality: High, Medium, Low) ANOVA was conducted with repeated measures on the last factor. No main effect of DC on W-S C-A was found $F(2, 112) = 1.29$, $p = .280$, $\eta_p^2 = .02$ and no interactions with DC were observed. Nevertheless, a main effect of MFT was observed $F(1, 56) = 4.24$, $p = .044$, $\eta_p^2 = .07$. Comparisons show that individuals who received MFT had a higher W-S C-A (*M* =.11, *SD* = .07) than those that did not (*M* = -.01, *SD* = .02) (see Table 41).

This finding suggests that receiving MFT allowed participants to apply higher confidence to correct decisions than incorrect decisions and/or lower confidence to incorrect responses. Further, the score went from a negative relationship to a small but positive relationship. Hence, , suggesting that they applied higher confidence to correct decisions and lower confidence to incorrect decisions. There was however, no main effect of playing games $F(1, 56) = 1.65$, $p = .204$, $\eta_p^2 = .03$ and no other interactions were observed, $p < .05$.

Table 41: Means and Standard Deviations for W-S C-A according to MFT, group and DC

MFT	Group	Overall DC	High DC	Medium DC	Low DC	Total
Yes	VGP	.05 (.17)	.13 (.30)	.10 (.33)	-.04 (.32)	
	NG	.13 (.22)	.25 (.27)	.08 (.38)	.14 (.41)	
	Total	.09 (.19)	.19 (.29)	.09 (.35)	.05 (.37)	.11 (.07)
No	VGP	-.03 (.16)	-.03 (.27)	-.02 (.33)	-.05 (.29)	
	NG	-.04 (.18)	.06 (.33)	.00 (.33)	-.00 (.49)	
	Total	-.04 (.17)	.01 (.30)	-.01 (.32)	-.03 (.40)	-.01 (.02)
Total	VGP	.01 (.17)	.05 (.29)	.04 (.33)	-.05 (.30)	
	NG	.05 (.21)	.15 (.31)	.04 (.35)	.07 (.45)	.05 (.06)
	Total	.03 (.19)	.10 (.30)	.04 (.34)	.01 (.38)	

Note. Standard deviations are in parenthesis.

9.6.4. Percentage Confidence in Correct and Incorrect Responses:

To further analyse the data, the percentage confidence in correct and incorrect decisions were examined. To do this the number of correct responses was recorded and the confidence in those decisions calculated to produce a confidence percentage in correct responses (see Table 42).

Percentage confidence in correct responses. A 2 (MFT: Yes, No) X 2 (Group: VGP, NG) ANOVA was conducted on percentage confidence data. A significant effect of MFT as found F(1,56) = 4.93, p = .031, η_p^2 = .08. Hence, participants who received MFT were found to be more confident in correct responses (*M* = 80.21, *SD* = 10.96) than those who did not receive the MFT (*M* = 73. 63, *SD* = 13.04). There was also a main effect of gaming F(1, 56) = 7.66, p = .008, η_p^2 = .12. Gamers (*M* = 81.01, *SD* = 10.19) were more confident in correct decisions than NGs (*M* = 72.81, SD = 13.19). No interaction was found between MFT and group F(1, 56) = .18, p = .672, η_p^2 = .003 was observed.

Percentage confidence in incorrect responses. Similarly, an ANOVA was also conducted on percentage in incorrect decisions. Unlike, confidence in correct responses, there were no significant difference in percentage confidence incorrect response in MFT F(1, 56) = 1.14, p = .290, η_p^2 = .02 However, participants who received feedback were more confident in their incorrect decisions (*M* = 77.63, *SD* = 10.28) compared to individuals who did not receive feedback (*M* = 74.39, *SD* = 14.19). Again, there were significant differences between VGPs and NGs F(1, 56) = 8.45, p = .005, η_p^2 = .13. VGPs were more confident (*M* = 80.42, *SD* = 10.06) than NG (*M* = 71.60, *SD* = 13.09). Again, no interaction F(1, 56) = .04, p = .851, η_p^2 = .001. In summary, no significant differences were found between percentage confidence in incorrect responses.

These findings demonstrate that participants who received MFT were more confident in correct responses. In other words, individuals who did not receive MFT remained a similar level of confidence in the decisions regardless of whether they were correct or incorrect. These findings suggest that MFT had a positive training effect. Interestingly, those who did not receive any feedback were slightly more confident in incorrect than correct responses. Importantly, although the MFT increased confidence in correct decisions, the W-S C-A is still relatively low.

Table 42: Means and Standard Deviations for % confidence (correct and incorrect) according to MFT and group

MFT	Group	% Correct	% Incorrect
Yes	VGP	83.68 (8.29)	82.33 (7.08)
	NG	76.73 (12.42)	72.93 (11.03)
	Total	80.21 (10.96)	77.63 (10.28)
No	VGP	78.36 (11.46)	78.51 (12.31)
	NG	68.89 (13.16)	70.27 (15.14)
	Total	76.92 (12.40)	74.39 (14.19)
Total	VGP	81.02 (10.19)	80.42 (10.06)
	NG	72.81 (13.19)	71.60 (13.09)
	Total	76.92 (12.40)	76.01 (12.40)

Note. Standard deviations are in parenthesis

9.6.5. WL and SA:

A 2 (MFT: Yes, No) X 2 (Group: VGP, NG) ANOVA was conducted on overall SA. No main effect of MFT on SA was observed F(1, 56) = 1.04, p = .313, η_p^2 = .02. Furthermore, no differences between VGP and NGs F(1.56) = .01, p = .938, η_p^2 = .00 and no interaction F(1, 56) = .025, p = .876, η_p^2 = .00. These results show that neither MFT nor group had an impact on SA scores (see Table 43).

A 2 (MFT: Yes, No) X 2 (Group: VGP, NG) ANOVA on reported WL scores was conducted. The results displayed no significant differences in WL between those who received feedback and those who did not $F(1, 56) = .28$, $p = .608$, $\eta_p^2 = .01$. This finding demonstrates that the additional instruction did not increase the subject feelings of overall WL. There was however, a close to significant differences between NG and VGPs $F(1, 56) = 3.84$, $p = .055$, $\eta_p^2 = .06$. Comparisons show that VGPs ($M = 54.53$, $SD = 14.52$) had lower WL than NGs ($M = 61.94$, $SD = 1.29$). No interaction between MFT and group were displayed $F(1, 56) =.11$, $p = .747$, $\eta_p^2 = .002$.

Table 43:Means and Standard Deviations for SA and WL as influenced by MFT and group

MFT	Group	Overall WL	Overall SA
Yes	VGP	52.93 (13.03)	18.80 (6.91)
	NG	61.55 (15.50)	18.40 (5.36)
	Total	57.24 (14.73)	18.60 (6.08)
No	VGP	56.13 (16.18)	16.80 (6.69)
	NG	62.30 (13.52)	16.93 (7.28)
	Total	59.22 (14.98)	16.87 (6.87)
Total	VGP	54.53 (14.53)	17.80 (6.76)
	NG	61.92 (14.29)	17.67 96.32)
	Total	58.23 (14.77)	17.73 (6.49)

Note. Standard Deviations are in parenthesis

9.6.6. SA Subscales:

The SA measure of SART is made up of attentional supply, demand and understanding. One way ANOVA were conducted on these subscales to examine the effect of group and MFT. The results showed that neither MFT nor group had an influence on feelings of any of the subsections of demand, supply or understanding SA. Due to the lack of significant findings, the results table reporting the means and standard deviation of SA scores as influenced by group and MFT are reported in Appendix 10e.

9.6.7. NASA TLX Subscales: There are 6 Subscales of Workload (mental demand, physical demand, temporal demand, performance, effort and frustration). To identify a specific type of WL one-way ANOVAs were conducted to examine these subscales by MFT and group (see Table 44).

9.6.7.1. Mental Demand: Results did not demonstrate an impact of MFT on mental demand $F(1, 56) = 2.39$, $p = .128$, $\eta_p^2 = .04$. However, there was a main effect of group $F(1, 56) = 9.85$, $p = .003$, $\eta_p^2 = .15$. NGs reported higher mental demand in the task ($M = 296.67$, $SD = 94.47$) than VGPs ($M = 217.83$, $SD = 101.18$). No interaction was observed $F(1, 56) = .32$, $p = .575$, $\eta_p^2 = .01$.

9.6.7.2. Performance: No impact of MFT on performance $F(1, 56) = .57$, $p = .453$, $\eta_p^2 = .01$ or gamer $F(1, 56) = .28$, $p = .601$, $\eta_p^2 = .01$. However an interaction between group and MFT was found $F(1, 56) = 5.20$, $p = .026$, $\eta_p^2 = .09$. NGs who received feedback reported less performance demands than VGPs. Suggests that the MFT reduced feelings of successfulness on the task. No other comparisons of the WL subscale were significant, $p < .05$.

Table 44: Means and Standard Deviations for dimensions of WL as influenced by DC and group

MFT	Group	Mental demand	Physical	Temporal Demand	Performance	Effort	Frustration
Yes	VGP	244.33 81.30	8.33 (24.32)	167.00 (127.85)	164.00 (104.73)	131.00 (90.32)	76.00 (62.08)
	NG	309.00 (90.75)	2.33 (4.95)	235.00 (120.48)	94.67 (51.49)	202.33 (89.72)	107.00 (118.44)
	Total	276.67 (90.82)	5.33 (17.52)	201.00 (126.86)	129.33 (88.42)	166.67 (95.60)	91.50 (94.24)
No	VGP	191.33 (114.42)	0.00 (0.00)	236.33 (126.95)	126.33 (96.72)	128.33 (73.69)	160.00 (132.30)
	NG	284.33 (99.62)	8.67 (28.25)	186.00 (102.96)	169.67 (116.92)	174.67 (91.66)	118.67 (125.55)
	Total	237.83 (115.53)	4.33 (20.12)	211.17 (116.42)	148.00 (107.71)	151.50 (85.05)	139.33 (128.46)
Total	VGP	217.83 (101.18)	4.17 (17.42)	201.67 (130.06)	145.17 (100.89)	129.67 (81.01)	118.00 (110.16)
	NG	296.67 (94.47)	5.50 (20.19)	210.50 (112.90)	132.17 (96.61)	188.50 (90.22)	112.83 (120.07)
	Total	257.25 (104.87)	4.83 (18.71)	206.08 (120.82)	138.67 (98.15)	159.08 (90.04)	115.42 (114.27)

Note. Standard deviations are in parenthesis

9.6.8. Between-Subjects Confidence-Accuracy:

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was also conducted. It was found that overall accuracy was not related to overall confidence $r(58) = -.02$, $p = .853$. A Bonferroni correction was applied for multiple comparisons, as such the level to reach significance was $p = .005$.

9.6.9. Relationship between WL, SA, and Accuracy, confidence and W-S C-A:

Confidence was found to be positively related to SA $r(58) = .47$, $p < .001$. The results demonstrate that higher decision confidence was related to higher reported SA scores. In addition, neither confidence nor accuracy was related to any other constructs.

Pearson's correlations were also conducted to establish whether accuracy, confidence, and W-S C-A were related to the psychometric scores. With the new significant level applied no significant findings were observed, however, there were some close to significant findings were reported. Decision accuracy negatively related to neuroticism $r(58) = -.27$, $p = .035$. This finding demonstrated that higher decision accuracy was related to lower scores of neuroticism. Suggesting that, individuals who score highly on the scale of neuroticism are less accurate in their decision making. On the other hand, decision accuracy was positively related to conscientious $r(58) =.27$, $p = .040$. Those who made more accurate decisions scored higher on the personality trait of conscientious. Decision accuracy was, however, closely and negatively related to Ambiguity $r(58) = -.29$, $p = .026$ and decision style $r(58) = -.30$, $p = .019$. Suggesting that being able to tolerate ambiguity is desirable to make more accurate decisions. No other relationships were found to be significant, $p > .005$.

9.6.10. Personality Constructs:

To assess and compare each group on the psychometric measures, one-way ANOVAs were performed on the data. Results showed that VGPs were significantly less conscientious than NGs $F(1, 59) = 6.02$, $p = .017$, $\eta_p^2 = .09$ No others findings were significant (see Table 45b).

Table 45: Means and Standard Deviations for dimensions of psychometric scores as influenced by group

a. Cognitive constructs

Gamer	Tolerance to Ambiguity	Decisiveness	Ambiguity Tolerance B	Decision Style
Yes	50.20 (8.73)	21.00 (4.49)	37.23 (6.40)	58.27 (10.20)
No	47.93 (6.51)	21.97 (5.75)	36.03 (6.45)	58.23 (10.87)
Total	49.07 (7.72)	21.48 (5.14)	36.63 (6.40)	58.25 (10.45)

Note. Standard deviations are in parenthesis

b. Personality

Gamer	Neuroticism	Extraversion	Openness To experience	Agreeableness	Conscientiousness
Yes	22.57 (8.34)	30.33 (7.37)	32.70 (7.92)	30.40 (6.12)	28.73 (6.64)
No	22.63 (8.48)	32.77 (6.02)	32.17 (6.25)	33.30 (5.56)	33.13 (7.23)
Total	22.60 (8.34)	31.55 (6.78)	32.43 (7.08)	31.85 (5.98)	30.93 (7.23)

Note. Standard deviations are in parenthesis

9.7. Discussion

The experiment presented in this Chapter examined the impact of MFT on decision accuracy, confidence and W-S C-A. The introduction of MFT to the task aimed to increase individual's metacognitive ability and to improve understanding into effective metacognitive training techniques. The findings from the experiment demonstrate a significant effect of MFT on W-S C-A. That is to say, individuals who received MFT were better able to discriminate between accurate and inaccurate response by applying higher levels of confidence in correct responses compared to incorrect responses. This is an important finding, as it demonstrates that metacognitive awareness can be improved via a small training element which has implications on recommendations for training initiatives. Other results from this experiment mirror that of the previous experiments. For example, both accuracy and confidence were influenced by DC and VGPs displayed higher confidence in their decisions, irrespective of their accuracy.

9.7.1. Accuracy:

In keeping with the previous experiments, individuals made more accurate decisions in high DC decision events. Similarly, no differences were found in accuracy between groups. Supporting the findings of the previous experiments that accuracy is impaired by low criticality. However, these results did not display any significant differences in accuracy with the addition of MFT. This could be explained by the fact that the MFT was designed to examine the relationship between confidence and accuracy and did not focus on accuracy specifically.

9.7.2. Confidence:

Furthermore, in relation to decision confidence, again, in support of the previous experiments, individuals were least confident in medium DC and showed higher confidence in their decisions in both high and low DC events. In contrast, the introduction of MFT did not influence confidence in the criticality of decisions made. The implications of this finding suggests that more direct training into the criticality of decisions to individuals which highlights the potential of the differences in confidence and accuracy in varying levels of DC. Replicating the findings from the previous findings, gamers in this task also displayed higher levels of confidence in their decisions.

9.7.3. W-S C-A:

In relation to W-S C-A score, both DC and group did not influence W-S C-A. However, significant differences were found in W-S C-A between those who received MFT and those who did not. The findings illustrate that individuals had a higher W-S C-A relationship if they had received the MFT prior to the task. As previously discussed, the direction of the relationship is not clear from this analysis alone; as such a positive relationship may indicate low confidence in incorrect responses as well as high confidence in correct responses. Further analysis was therefore conducted on percentage confidence in correct and incorrect responses. Results found that individuals who had received the MFT were able to apply more confidence to correct decisions than incorrect decisions in comparison to individuals who did not receive feedback. The results imply that individuals who did not receive the MFT displayed similar levels of confidence in their decisions regardless of if they were correct or incorrect. Interestingly, in this group, a slightly higher confidence was displayed in incorrect decisions. Previous literature on feedback and training, which aimed to improve metacognitive awareness by reducing confidence, have found strong evidence for a positive influence on reducing bias in decision making (Kim, 2018; Fiorella et al., 2012). Specifically, literature has suggested that improving an individual's awareness of their thinking processes can help improve decision making and metacognition (Kim, 2018; Cohen et al., 1998).

It is important to note that, although W-S C-A improved, decision confidence remained high regardless of whether the individuals received MFT, with a general tendency to overconfidence. As such, further work would be needed to further improve the W-S C-A relationship taking into account such aspects. One limitation of the experiment was that it was conducted on novices; none of the participants had any prior air defence experience. Hence, the low metacognitive skill improvement could be explained by the lack of familiarity for the task as previously noted and discussed in Chapter eleven.

Moreover, studies have shown that people do not apply their training to unfamiliar and dissimilar domains because they lack the necessary metacognitive strategies to recognise the underlying problem structure (Barnett & Ceci, 2002), thus, and as suggested by the previous research, individuals were unaware of their poor performance in the task. Hence, raising individual awareness of their abilities would be vital to future training success.

Nevertheless, these are promising results for a simple MFT technique which was inexpensive and did not require time or expertise. MFT in this study was provided by short instruction and PowerPoint, combined with a practice trial and direct personal guidance on their performance. As discussed, other techniques have been successful in de-biasing and reducing over confidence and used for training. These include video games and paper instructions (Kim, 2018; Morewedge et al., 2015). Hence, these results may be useful in moving forward and designing MFT.

9.7.4. WL and SA:

Moving on, performance measures of SA and WL were also included. Previous research has demonstrated that SA can be improved via critical thinking training skills, which help metacognitive ability (Cohen et al., 1998). However, in the current experiment, MFT had no effect on SA and suggests that the inclusion of MFT by the use of the technique chosen did not increase individual SA. As previously mentioned the MFT provided to individuals in this experiment was specifically related to an individual's awareness of the accuracy of their response and was not tailored towards improving SA. Hence, future work could examine the types of training needed which focuses on different aspects of decision making.

It is also important to be aware that providing individuals with additional information may have an impact on feelings of WL, thus it is important than any intervention does not provide an additional load to the decision maker. As such, the experiment was also interested in the impact of WL. Individuals who received MFT did not report higher levels of WL in comparison to those who had no MFT training and thus demonstrating a positive effect of training effect. This supports the previous finding in the other experiments in the thesis as W-S C-A was found to be unrelated to WL. An explanation could, therefore, suggest that WL and W-S C-A rely on different cognitive resources. WL is related to an individual's attentional resources and is subjective to the individual. On the other hand W-S C-A is based both on subjective and objective metrics. Furthermore, individuals may be more sensitive to changes in WL in comparison to their performance in the task. Future work could expand on the cognitive process and the relationship between WL and W-S C-A.

The results from this experiment also found differences in the group's feelings of WL. NGs reported more mental demand on the NASA TLX subscale. As discussed, VGP has been linked to higher cognitive ability and inability to deal with WL has been linked to lower cognitive ability (Gonzalez, 2004). Furthermore, this experiment demonstrated an interaction between MFT and group in the subscale of performance. NGs who received feedback reported fewer performance demands than VGPs, implying that the MFT actually reduced feelings of successfulness on the task for NGs.

9.7.5. Personality and Cognitive Constructs:

In line with the previous experiments, measures of personality and cognitive constructs were also collected to examine the interplay between the constructs. The findings suggest that there are personality traits which may relate to the accuracy of decisions. In this study, close to significant findings were reported and decision accuracy was closely and negatively related to neuroticism, replicating the findings from the third experiment. Trends in the results suggest that higher decision accuracy was related to lower scores of neuroticism. Individuals who score high on neuroticism tend to be more anxious and worry more. Hence, this negatively impacted on participants' performance in this task. Indeed, research has shown that cognition is impaired in highly neurotic individuals (Bryne et al., 2015). These findings suggest that there may be scope to examine individual traits and the impact specific traits have on performance in air defence. Decision accuracy was also positively related to the personality trait of conscientious. Those who made more accurate decisions scored higher on measures of conscientiousness. Individuals who score higher on the scale of consciousness are regarded as being more thoughtful and have a desire to perform well in the task (Costa & McCrae, 1995). Subsequently, the results presented here suggest that the personality traits of low neuroticism and high conscientious positively relate to decision making in this task. Although the traits in this study were not found to relate to MFT, previous research has demonstrated these traits to be predictive of training success as well as in military decision making (Saus et al., 2012) they may, therefore, also be relevant to the air defence domain. Future research could examine more specific training targeted at increasing metacognitive ability.

In terms of measures of cognitive construct, the current experiment found additional support for the findings of Experiment 1 and 2. Decision accuracy was negatively related to ambiguity and decision style which again supports the notion that these cognitive constructs may be advantageous for individuals in complex environments in order to perform well. However, although, previous research has shown that certain individual differences are beneficial for successful training performance (Saus et al., 2012) the results from this study failed to find a relationship between any of the personality traits and cognitive constructs.

9.8. Summary

In summary, this experiment found that MFT had a positive effect by lowering confidence in incorrect decisions. However, more work would be required to decrease the tendency for overconfidence and to help improve the W-S C-A relationship, as well as the need to conduct this work with experts. Individual differences have again shown to be influential in decision making and in particular there is support to tolerance to uncertainty as a positive tendency that is required to make accurate decisions.

CHAPTER TEN

10. Investigating the Impact of DC and TL on Decision accuracy, confidence and W-S C-A in Principal Warfare Officers (PWOs)

10.1. Introduction

As previously discussed, experts in comparison to novices will rely on different decision making strategies and tend to base their decision on previous experience (Klein et al., 1986). To increase the validity of the previous findings, build upon and integrate experimental work into NDM research domain, it was important to conduct research on domain experts (Johnston et al., 1998). In this experiment, PWOs who had recently finished their PWO training course were recruited to take part. As mentioned earlier (see Chapter 1), PWOs are the main decision makers in an Ops room and it is thereby important that decisions taken are correct and the appropriate level of confidence is applied to those decisions. Furthermore, it is important to understand how decisions are made with individuals who are experienced in making similar decisions. To assess the impact of DC, TL and individual differences on decision making, the first experimental set up was replicated with expert RN PWO participants.

10.2. Hypotheses

- **HP18:** High DC will reduce decision accuracy (A), confidence (B) and W-S C-A (C).
- **HP19**: High TL will reduce decision accuracy (A), confidence (-B) and W-S C-A (C).
- **HP20:** Individual differences will be found in relation to decision accuracy, confidence and W-S C-A.
- **HP21:** To explore the relationship between accuracy (A), confidence (B), W-S C-A (C) and measurements of WL and SA (D).

10.3. Participants

Twenty-two PWOs were recruited. Defence Council Instructions (DCI) and Temporary Memorandums were used to target the correct level of decision maker for the research and the recruitment of experts was facilitated by Dstl as described in section (4.3.3.2). The mean age of participants was 33 years old (*SD* = 2.62) and the length of time spent in the RN ranged from 3.5 years to 20 years (*M* = 9.00). Ethical approval and statistical criteria remained the same for this experiment (see Section 4.3.2 and 5.3).

10.4. Results

To assess the differences in means a number of statistical analyses were performed on the data for accuracy, confidence and W-S C-A using ANOVAs. A manipulation check was carried out to assess the differences in TL (see analysis of Workload and Situational Awareness). No significant differences were found in WL, $p = .474$ and SA, $p = .707$, consequently, the TL manipulation not successful. An alpha level of .05 was used for all statistical tests unless otherwise stated.

10.4.1. Accuracy:

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data (see Table 46). A main effect of DC on decision accuracy was found $F(2, 38) = 23.77$, $p < .001$, $\eta_p^2 = .67$. PWOs were more accurate in high DC decisions (*M* = 5.59, *SD* = 2.09) than low DC (*M* = 2.59, *SD* = 1.71), $p = .001$. Individuals were also more accurate in medium DC (*M* = 6.27, *SD* = 1.68) than low, $p < .001$. Overall, PWOs in this task were most accurate in the medium DC decisions. However, no main effect of TL on decision accuracy $F(2, 38) = .14$, $p = .896$, $\eta_p^2 = .02$ was observed. There was also no interaction shown between DC and TL $F(4, 38) = .27$, $p = .893$, $\eta_p^2 = .03$. As a result these findings demonstrate that PWOs made the most accurate decisions when presented with decisions with medium levels of criticality and made more incorrect responses when presented with low DC decisions.

Table 46:Means and Standard Deviations for Accuracy as influenced by DC and TL

TL	Overall Accuracy	High DC	Medium DC	Low DC
High	15.00 (3.37)	5.57 (2.64)	6.71 (1.70)	2.71 (1.90)
Mod	14.14 (2.73)	5.71 (2.29)	6.29 (1.70)	2.14 (1.07)
Low	13.88 (3.52)	5.50 (1.60)	5.88 (1.73)	2.88 (2.17)
Total	14.32 (3.12)	5.59 (2.09)	6.27 (1.68)	2.59 (1.71)

Note. Standard deviations are in parenthesis.

10.4.2. Confidence:

A 3 (Task load [TL]: High, Moderate, Low) X 3 (Decision Criticality [DC]: High, Medium, Low) mixed ANOVA, with repeated measures on the last factor, was conducted on the data. There was a main effect of DC on decision confidence F(2, 38) = 4.29, p = .021, η_p^2 = .18. Thus, PWOs were more confident in low DC (M = 45.86, SD = 3.85) than medium DC (M = 43.86, SD = 4.93), p = .041. However, again no main effect of TL on confidence F(2,19) = .09, p = .909, η_p^2 = .01. Nevertheless, there was an interaction between DC and TL F(4,38) = 3.19, p = .024, η_p^2 = .25. By examining the descriptive statistics for trends, in the low TL, both high and low DC decisions had higher confidence levels applied. Comparatively, in the high TL, PWOs had equal levels of confidence across low and medium DC. The moderate TL created the lowest decision confidence when making medium DC decisions. Low DC varied the least (see Figure 16). These results show that DC also impacted on decision confidence. However, PWOs were more confident in low DC than both medium and high DC. TL manipulation in this task did not influence PWOs decision confidence.

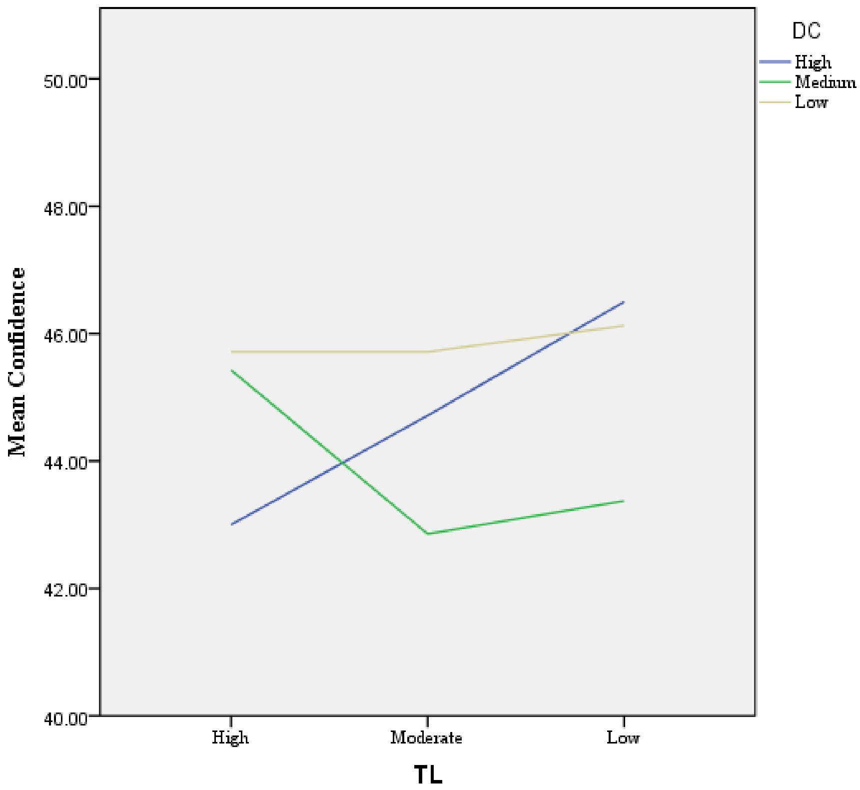


Figure 16. Graph showing the interaction between DC and TL in PWO participants

Table 47: Means and Standard Deviations for Confidence as influenced by DC and TL

TL	Overall Confidence	High DC	Medium DC	Low DC
High	134.14 (14.87)	43.00 (5.66)	45.43 (4.86)	45.71 (4.96)
Mod	133.29 (9.48)	44.71 (2.93)	42.86 (5.11)	45.71 (3.04)
Low	136.00 (12.01)	46.50 (3.96)	43.38 (5.15)	46.13 (3.91)
Total	134.55 (11.77)	44.82 (4.36)	43.86 (4.93)	45.86 (3.85)

Note. Standard deviations are in parenthesis

10.4.3. W-S C-A:

A 3 X 3 mixed ANOVA was performed on the relationship between TL, DC on individuals W-S C-A. A main effect of DC on W-S C-A was found

F(2, 38) = 4.02, p = .022, η_p^2 = .81. However, comparisons only show a close to significance level, p = .078. As such, PWO's W-S C-A was higher in the high DC category (M = .18, SD = .39) compared to low DC (M = -.21, SD = .49). No interaction between TL and DC F(4,38) = .30, p = .879, η_p^2 = .03 and no main effect of TL F(2, 19) = .12, p = .892, η_p^2 = .01.

These findings suggest that PWOs are able to distinguish between accurate and inaccurate responses when presented with high DC events. In contrast, TL had no impact on metacognitive ability in this task in the PWO population.

Table 48: Means and Standard Deviations of W-S C-A as influenced by DC and TL

TL	Overall W-S C-A	High DC	Medium DC	Low DC
High	-.01 (.19)	.15 (.47)	.05 (.43)	-.14 (.45)
Mod	-.06 (.30)	.26 (.35)	-.01 (.32)	-.34 (.57)
Low	.04 (.21)	.14 (.41)	.02 (.27)	-.14 (.47)
Total	-.01 (.23)	.18 (.39)	.02 (.33)	-.21 (.49)

Note. Standard deviations are in parenthesis.

10.4.4. Percentage Confidence in Correct and Incorrect Responses:

W-S C-A demonstrates the relationship between confidence and accuracy. However, a high correlation suggests both being highly confident in correct decisions as well as low confidence in incorrect decisions. Similarly, a negative correlation would suggest that individuals are highly confident in incorrect responses or not confident in correct responses. By examining the percentage confidence in incorrect or correct responses, the direction of the confidence (over/under confidence) can be displayed. No significant differences on percentage confidence in TL in correct responses

F(2,21) = .24, p = .790, η_p^2 = .03 was found and no main effect of TL was shown for incorrect decisions F(2,21) = .02, p = . 984, η_p^2 = .00.

Table 49: Means and Standard Deviations for % confidence (correct and incorrect) according to TL

TL	% Correct	% Incorrect
High	89.27 (10.30)	90.37 (10.57)
Mod	87.95 (8.21)	89.83 (5.64)
Low	91.16 (8.55)	90.59 (8.01)
Total	89.54 (8.71)	90.28 (7.90)

Note. Standard deviations are in parenthesis.

10.4.5. WL and SA:

No significant differences in WL F(2, 19) = .78, p = .474, η_p^2 = .08 or in SA F(2,19) = .35, p = .707, η_p^2 = .04. This finding suggests that the TL manipulation was not successful. Furthermore, none of the WL and SA subsections were found to be significant.

Table 50: Means and Standard Deviations for WL and SA as influenced by DC and TL

TL	Overall WL	Overall SA
High	38.40 (25.76)	26.29 (8.90)
Mod	28.96 (18.79)	23.57 (6.95)
Low	26.93 (8.98)	22.88 (8.51)
Total	31.22 (18.53)	24.18 (7.93)

Note. Standard deviations are in parenthesis.

10.4.6. Between-Subjects Confidence-Accuracy:

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was also conducted. It was found that overall accuracy was not related to overall confidence $r(20) = -.06$, $p = .784$.

10.4.7. Relationship between WL, SA and Accuracy, confidence and W-S C-A:

To assess the relationship between WL and SA a series of Pearson's correlations were calculated. A Bonferroni correction was applied. No significant findings between WL, SA, Accuracy, confidence or W-S C-A were found, $p > .05$.

Table 51: Means and Standard Deviations for dimensions of psychometric scores

Measure	PWO
Tolerance to Ambiguity	48.86 (5.83)
Decisiveness	18.36 (5.22)
Ambiguity Tolerance B	34.68 (3.93)
Decision Style	52.64 (6.59)
Neuroticism	19.45 (7.41)
Extraversion	31.95 (6.46)
Openness To experience	30.36 (6.36)
Agreeableness	27.32 (3.12)
Conscientiousness	32.23 (5.71)

Note. Standard deviations are in parenthesis.

10.5. Discussion

The experiment reported in this Chapter was interested in the application of the experimental method to an expert population to assess the ecological validity of the previous findings, as well as NDM criteria set out by Johnston et al. (1998). The experts in this task were made up of current RN PWOs. In accordance with the first experiment, DC and TL were manipulated to examine the impact of decision accuracy, confidence and W-S C-A. Individual differences in PWOs were also examined. The results from this experiment replicate some of the findings of those reported in the novice populations. In support of the previous findings, PWOs were also found to be more accurate in the higher criticalities compared to the low DC, thus, increasing the ecological validity of these findings. Additionally, TL was also found not to have an influence on decision accuracy, confidence or W-S C-A. However, in this experiment, WL and SA were unaffected by the TL conditions which has implications on the experimental design for use in expert populations. Further, no individual differences in personality and cognitive constructs were observed.

10.5.1. Accuracy:

DC was found to influence decision accuracy in the task. PWOs were more accurate in the higher criticalities. Thus, supporting previous finding and literature that criticality plays a role in performance (Hanson et al., 2014; Callister et al., 1999). However, unlike the previous experiments which found high DC events to be most accurate, PWOs were slightly more accurate in the medium DC. Hence, in line with the previous studies, it is proposed that higher DC events were deemed more important by the participants and therefore performance increased in these decisions. Further, the finding that medium DC was highest in this population could suggest that these decisions were more familiar/easier to make and these levels of decisions are what the experts are used to. Experts used prior experience to make decisions (Klein et al., 1986**)** which in this task assisted them in making more accurate decisions. The findings, therefore, imply that criticality influences performance and cognition is not impaired in these decision events. This is an important finding for air defence decision making. On the other hand, however, caution should be applied to decisions which are low in criticality as these results suggest that it is more likely that an inaccurate decision would be made.

10.5.2. Confidence:

The findings from decision confidence suggest that, although PWOs were more accurate in medium DC events, these decisions were associated with the lowest decision confidence. As previously discussed, it has been suggested that the medium DC produces the greatest uncertainty which has been linked to reduced confidence (Heerman & Walla, 2011). This is a particularly important finding which has implications for information seeking behaviour, as confidence predicts information seeking (Desender et al., 2018). The authors found that when participants were low in confidence they tend to seek further information. This occurred regardless of the accuracy associated with the decision. In an air defence domain, this could have critical consequences in terms of timely decision making, as decision makers may feel the need to gather more information to make a decision if not correctly associated with the accurac of decisions. Thus, these findings provide further evidence that metacognitive abilities are a desirable skill in air defence operators.

In comparison, the low DC events induced greater feelings of confidence in decisions. Nevertheless, a possible explanation for this finding could be linked to the nature of the task as experience and familiarity have also been liked to confidence (Wheatcroft et al., 2017). Thus, as the PWOs have a greater experience level in making similar decisions, they held an incorrect belief in the performance of the task. This finding has clear implications to air defence decision making as overconfidence has been linked to riskier decision making and closing down of hypothesis generation (Yang et al., 2012). Another possible explanation could be that individuals felt the low DC events were easy decisions to take and this gave them a sense of greater confidence in their responses. Furthermore, future training should consider the impact of perceived task difficulty in the design of experimental work should be taken into consideration. Furthermore, similarly to the previous experiments, conducted with novices, percentage confidence remained high and PWOs were 90% confident in incorrect responses. That is to say that regardless of the accuracy of the decision, confidence in the decision remained high. This was disproportionate to the accuracy of the decisions. The findings also highlight the importance of understanding the relationship between accuracy and confidence because although PWOs were the most accurate in medium decision events, they were also the least confident.

10.5.3. W-S C-A:

The results from W-S C-A found a close to significant impact of DC on W-S C-A. PWOs demonstrated slightly higher W-S C-A scores in high DC which would suggest a better awareness for these types of decision events. Again, this could be explained by expert experience in more challenging decisions, that they are deemed more important and that these decisions are made with the correct amount of confidence. Furthermore, experience impact on metacognition (Lichtenstein et al., 1977) these events may be more familiar to expert decision makers. These outcomes also support the findings of the second experiment, in which W-S C-A was higher in high DC events. Hence, there is strong support for individual metacognitive ability improving in response to higher levels of criticality. In contrast, low DC was found to have the lowest W-S C-A, this might be explained by individuals being less accurate but more confident. Overall, the low W-S C-A scores imply that individuals seem to be unaware of when they are incorrect. Consequently, a lack of awareness of the decision accuracy may have further implication in the air defence domain.

10.5.4. Personality and Cognitive constructs:

In regards to the individual differences, no personality traits or cognitive constructs found to be significantly related to decision accuracy, confidence, W-S C-A, SA or WL. Of course, the small sample size in this experiment could explain the lack of significance in these findings.

As well as the limitation acknowledged by the sample size, the PWO participants had also only recently qualified as PWOs thus, only newly qualified for the role. Although on average they had 9 years' experience, it is envisioned that there may be differences in the level of experiences of the participants with PWOs who had more extensive experience in the role. Further work could, therefore, consider more experienced PWOs.

10.6. Summary

The findings of this experiment support DC as an important consideration to examine in the investigation of decision making in critical environments, particularly air defence. Hence, the criticality of a decision has important implications on decision making. Decision accuracy and confidence was found to be influenced by DC with more accurate decisions being made when presented with decisions of higher criticality. Furthermore, PWOs displayed higher metacognitive awareness when presented with more critical decisions. This finding suggests that cognition may not necessarily be impaired in these decisions, however; a low metacognitive awareness in low critical decisions may be problematic.

The replication of results from novice participants provides ecological validity to the findings presented previously in the thesis. In addition, it also presents the first piece of research that has used this method on serving RN experts. Consequently, the use of experts in this experiment provided valuable insight into the design and conduct of experimental work with experts. However, as discussed in Chapter three, gaining access to a range of experts in the air defence domain is difficult. Consequently, the expert sample size in this experiment is a potential limitation and future work could address this issue. Furthermore, future research would need to ensure the task manipulations are the most suitable for use with experts, whilst retaining the balance of ecological validity and research endeavour (see Chapter 12 for a more detailed discussion on limitations and future work).

PART THREE: Discussion and Implications of Findings

CHAPTER ELEVEN:

11. General Discussion

11.1. Introduction

Chapter eleven discusses the findings in the Experimental Work carried out with respect to the aims and research questions outlined at the start of this thesis and the existing literature. To reiterate, four research questions were asked in the current thesis:

1) What are the external factors that influence decision confidence and accuracy in an air defence decision making task? Specifically, TL, DC, time pressure and audio communications, and what relationships exist with metacognition?

2) What, if any, are the individual differences involved in air defence decision making and how do they relate to metacognitive skills. For example, in light of personality constructs, cognition and video game play?

3) How does the method and measurements used in this thesis align with the wider methods and measurements currently used to assess performance in decision making?

4) Can metacognitive ability in air defence be improved through feedback training? In light of these questions, the current Chapter will address the findings.

11.2. Research Question 1: What are the external factors that influence decision accuracy and confidence in an air defence decision making task?

A key aim of the Experimental Work was to examine a range of external factors in relation to decision accuracy, confidence and metacognitive ability in air defence decision making. External factors referred to those factors over which the operator has no control. To achieve this, experimental techniques akin to classical DM were adopted. TL, DC, time pressure and audio communication were manipulated as part of the experimental studies. The following sections will discuss the findings of the impact of these external factors in relation to accuracy, confidence and W-S C-A in turn.

11.2.1. Accuracy:

It was hypothesised that decision accuracy would be influenced by DC and TL, and that differences would be observed between different groups. It was stated that high DC would reduce decision accuracy. This hypothesis was not supported. It is of note however, that all of the findings across each experiment found that high DC impacted positively on the performance of individuals. It has been previously stated that criticality is important to decision making, but DC thus far has been ill-defined and researched (Hanson et al., 2014). The definition used in the current thesis expands upon the definition provided by Hanson et al. (2014) to include a range of criticalities, with a focus on the concept that each decision event in a given scenario can be associated with a level of criticality, and not merely a scenario as a whole.

The results contained in this thesis indicate that individuals were consistently more accurate in decisions that were associated with either high or medium criticality. Thus, individuals made fewer errors in decisions that held higher consequences, and such outcomes lend support to the findings of Adams-White et al. (2018). The finding that individuals made more accurate decisions in more critical decisions was, therefore, a consistent and stable outcome. This was found to be true regardless of the other factors such as time pressure (see Chapter 6), audio (see Chapter 8) and metacognitive feedback (see Chapter 9). It is a finding that was further replicated by experts (see Chapter 10). Hence, it can be concluded that high criticality decisions did not impair performance in the task as hypothesised. One possible explanation for this is the assumption that task performance increases when participants find the task more important (Kliegel, et al., 2004). Hence, it could be that more critical decisions were deemed more important. Further, these findings support, Harris et al. (1995) who found that accuracy increased after an exposure to a critical task. Research has also demonstrated that critical tasks do not actually impair cognitive ability (Callister et al., 1999). With this in mind, it is possible that this may have created conditions which allowed individuals to answer correctly. Consequently, criticality has been found to have a positive impact on performance. Yet, research has also shown high criticality to negatively impact on accuracy (Hanson et al., 2014). Nevertheless, this latter finding was not supported by the research reported in this thesis. Subsequently, the results found here also highlight the increased occurrence of errors in decisions of lower criticality. Low DC events held less consequence if an incorrect decision was made. However, it is still important that accurate decision are made, as errors in these types of decisions may lead to an incorrect assessment of the situation, particularly as they are generally made earlier in a given scenario. As a result, one key message is that, individuals need to be more aware of the likelihood of making errors when faced with less critical decisions in air defence.

A key manipulation across the Experimental Work was that of TL. It was hypothesised that TL would reduce decision accuracy. TL was varied, and participants were randomly allocated to either a high, moderate or low TL condition. The aim was to examine whether the differences in cognitive load and temporal demand, (i.e. whether TL) had an impact on decision accuracy, confidence and metacognitive ability. Indeed, TL was found to influence decision accuracy. The first experiment found that individuals were more accurate in the low TL condition compared to the moderate condition. As such, this finding would suggest moderate TL decreased performance in the task. In contrast, a similar level of accuracy was observed between both high and low TL conditions. This finding is contrary to the Yerkes-Dobson law (1908) in which it would be expected that individuals would perform better in the moderate TL condition. However, this finding was not replicated in subsequent experiments, as no differences in accuracy between TL conditions were observed. The finding that accuracy did not differ in the TL conditions is a valuable finding as differences were found in reported feelings of WL between the conditions. To recall, measures of WL were also collected during the tasks which assessed individuals' feelings of subjective WL. The findings from WL showed higher TL conditions increased individual's feelings of WL during the task. However, as no differences were found in TL, the results may, therefore, imply that TL influences mental and cognitive performances but does not necessarily impact on overall performance in the task. This finding also suggests that TL and WL may impact individuals differently. Findings from the manipulation of WL are discussed in more detail later

The Experimental Work also considered external factors of time pressure (see Chapters 6 & 7) and audio communication (see Chapter 8). The results from the second experiment, which reduced the time for participants to make a decision, mirrored those of the first experiment. The results demonstrated that decisions were made with higher accuracy in the higher DC events. Furthermore, decision accuracy was also not influenced by TL in these conditions. To investigate whether there were any differences between the time to make a decision between 10 seconds and 20 seconds, further analysis was conducted. Comparisons between the two time conditions (see Chapter 7) found no differences between decision accuracy. This finding illustrates that reducing the time to make a decision did not negatively impede performance. This outcome supports the research of Yang et al. (2012) who similarly found no effect of time pressure on accuracy. However, it is, of course, possible that the time pressure manipulation in the experiment conducted in the current thesis may not have been strong enough, due to the lack of significance reported in temporal demand subscales of WL. Further research to examine what contributes to time pressure in relevant decision making tasks is warranted. In addition, Chapter eight reported upon the introduction of an audio element to the experimental conditions. The results showed that audio that was both attended to and not attend to did not influence the accuracy of decisions. This finding supports Beaman et al. (2014) who found that interruptions, such as the presence of audio, do not necessarily impact on individual's decision accuracy. Hence, neither time pressure nor audio impacted on the accuracy of decisions.

The results from the Experimental Work and the investigation of the external factors that influence decision accuracy would imply that DC is one of the main contributors to performance during a decision making task. Highly critical decisions are made with more accuracy whilst accuracy was not impacted by external factors of TL, time pressure, or audio. It is important to note that decision accuracy in the task was scored against pre-approved SME 'correct/best decision' responses. However, accuracy scores were generally low. This suggests that the task set was perhaps too difficult. Potential reasons for the lack of support for the hypotheses are discussed in Chapter twelve. However, it is not only the accuracy of a decision that is important, but also the corresponding confidence.

11.2.2. Confidence:

It was hypothesised that decision confidence would be influenced by DC and TL and that differences would be observed in the different groups. Specifically, it was hypothesised that DC would reduce decision confidence. The results showed that individuals tended to be less confident in their decisions when responding to decision events of medium criticality. An explanation for this might be that the medium criticality decisions created higher levels of uncertainty and confusion, thus causing a decrease in confidence levels. This supports previous findings that uncertainty can decrease decision confidence (Heerman & Walla, 2011). In addition, medium DC may have been regarded as more difficult, thus supporting the findings of Kebbell et al. (1996). This is an interesting finding, as in comparison to confidence, as previously described, the medium DC decisions were generally made with more accuracy. Such an outcome would imply that, although individuals were more accurate in their decision, this was not reflected in their decision confidence. It could, therefore, be argued that a medium level of criticality impairs individual's metacognitive ability which will be discussed later on. Hence; decision confidence was most impaired by decisions of medium criticality and individuals displayed higher levels of confidence in the low DC decisions.

It was also predicted that TL would reduce decision confidence. On the contrary, the results found that TL did not impact on the confidence of decisions. Participants remained reasonably equivalent across the different TL conditions in respect of reported confidence. This finding could be explained by confidence being relatively stable and uninfluenced by the demands of the task (Pallier et al., 2002; Burns et al., 2016). However, as previously discussed, confidence was found to be influenced by DC. This finding implies that further research is needed to investigate the types of decisions to be made within a critical environment and provides further support that the relationship between confidence and accuracy (i.e., W-S C-A) warrants investigation.

Furthermore, in line with the findings from decision accuracy, neither time pressure nor audio impacted on decision confidence. However, again this finding could be related to the lack of success of the task manipulations. The results of investigating the impact of external factors on decision confidence demonstrate that DC influences confidence. Higher confidence was placed in low DC events. However, the external factors of TL, time pressure and audio were not found to influence decision confidence. The next section will focus on the relationship between confidence and accuracy, as assessed by the W-S C-A metric.

11.2.3. W-S C-A:

As expressed previously, one main focus of the thesis was the relationship between accuracy and confidence, as it has crucial implications for future events and decisions taken. This thesis argues that it is important that individuals are aware of their thought processes (i.e., metacognition). In the current thesis, metacognitive ability was assessed by the relationship between confidence and accuracy as measured by W-S C-A. To reiterate, in the experiments conducted, once a decision had been made, individuals were required to provide a confidence rating in relation to that decision. A high confidence score indicated that they were extremely confident that they had made the correct decision in the given situation. The findings for W-S C-A however, were varied, with only one study finding any significant differences. In the second experiment, W-S C-A was influenced by the criticality of the decision. The results showed that individuals had a slightly higher W-S C-A in high DC decisions. Moreover, this trend was also shown in the expert population (see Chapter 10) suggesting that, individuals were better able to discriminate between correct and incorrect responses when making decisions that were high in criticality. As previously noted, the higher accuracy scores in highly critical decisions could be explained by decision makers deeming these to be more important (Kliegel et al., 2004). Hence, an explanation for this finding is that the higher W-S C-A is derived from individuals having to apply themselves more to that decision which leads to an increased ability to apply the appropriate confidence level. This provided further support research which has suggested that criticality does not impair cognition (Callister et al., 1999).

However, the finding that W-S C-A was significantly impacted by criticality was not replicated in any of the other experiments reported in this thesis. Contrary to expectations, but in agreement with Adams-White et al. (2018), none of the other external manipulations had an impact on an individual's ability to discriminate between accurate and inaccurate responses. Furthermore, in the reported experiments, the overall correlations between confidence and accuracy were either low or close to zero. As previously mentioned, the absence of a relationship between confidence and accuracy could be explained by the lack of experience of the novice population in this study, particularly as training and experience was shown to increase calibration (Lichstenstein et al., 1977).

The results in the thesis provide further support to the idea that the cause of mis-calibration is a tendency to apply consistent confidence levels irrespective of their accuracy level (Pallier et al., 2002). This could suggest that confidence may be a stable trait and not related to an individual's awareness of their accuracy; that is to say, irrespective of accuracy, individuals express a consistent confidence level (Stankov et al., 2015). This is further supported by the findings which indicate that confidence is stable across TL conditions and across the experiments. Indeed, the results from the experiments indicate that, regardless of the load on the operator, confidence is largely unaffected. One explanation for this is that there is a general confidence factor (Kleitman & Stankov, 2007) and individuals have a habitual way in which they assess the accuracy of their decisions (Stankov et al., 2015). This is supported by the findings in this study, where no significant differences were found between confidence scores in the TL conditions.

As well as examining the within-subjects relationship, the relationship between overall accuracy and confidence was assessed (i.e., between-subjects). No relationship was found between decision accuracy and confidence. Though some research has shown significant relationships between confidence and accuracy (see Kebbell et al., 1996), the results also support the notion that has been repeated in other domains, whereby confidence is not a good indicator or predictor of accuracy. This has implications on group decision making, as individuals may base their own decisions on others' confidence. In the group dynamics of a Ship's Ops room, this should be taken into consideration and would certainly warrant further research (see Future Work for a discussion).

To summarise the findings into external factors that influence decision accuracy and confidence, a main and consistent finding from the Experimental Work was that the criticality of the decision played a crucial role in both decision accuracy and decision confidence, with some findings also relating DC to metacognition. Hence, the research contained in the thesis has built on the knowledge regarding how criticality is defined and the impact on decision making in relation to accuracy and confidence in critical environments. As discussed in Chapter one, air defence decision making involves a range of tasks. Some of these decisions will be low in criticality such as Routine tasks. In comparison, Threat Response tasks require a response to an impending threat. As such, these decisions are associated with a much higher criticality. Consequently, it can be argued that criticality plays an important role in decision making. The finding that DC impacted on both performance and confidence is an interesting finding, given that individuals were not informed about the criticality of decisions before the task. This contrasts previous experiments in which participants were informed about the criticality of the scenario prior to the task (Bliss & McAbee, 1995; Hanson et al., 2014). Overall, the research contained in this thesis has found important findings surrounding the criticality of the decisions. It has been demonstrated that more errors in decision making are made when the decision criticality is low and the decision is made in a moderate TL condition. Further, decision confidence is impaired in medium DC events. There is also some evidence to suggest that individuals are better able to discriminate between correct and incorrect responses for highly critical decisions. Fundamentally, research into DC is warranted.

11.3. Research Question 2: What, if any, are the individual differences in air defence decision making and how do they relate to metacognitive skills?

As well as understanding the influence of external factors, the current thesis aimed to examine individual differences in decision making in critical environments and their impact on decision accuracy, confidence and metacognition. Individuals vary in their ability to estimate the uncertainty and reliability of their choices as well as their ability to estimate the reliability of their own decisions (Fleming et al., 2010; Song et al., 2010; Wheatcroft et al., 2017). The experiments in the current thesis examined individual differences in the light of broad personality traits (Extraversion, Neuroticism, Openness to Experience, Agreeableness & Conscientiousness- Costa & McCrae, 1992) and cognitive constructs (Tolerance to Ambiguity, Decision style - Budner, 1961; Roets & Van Hiel, 2007) to investigate how these individual differences might influence decision confidence and decision accuracy. In addition, the research examined individual differences with regards to different populations. Previous research found that VGPs may be better suited to certain decision making tasks (Wheatcroft et al., 2017). As yet, however, there has been little research on how VGPs perform in air defence decision making environments. To recall, the participants in the experiments were divided into three groups, one non-gamer (NG) group and two gaming groups. The latter were selected depending on how many hours of video games they played (i.e., VGP1 = *less than 7 hours per week*, VGP2 = *more than 7 hours per week*). By investigating different groups, the research aimed to gain insight into the development of expertise and how skills transfer into other domains. Hence, this thesis also aimed to highlight the potential benefits of using different populations to examine decision making. Subsequently, the findings in the thesis may be beneficial in building research into the possible and relevant person characteristics suitable for air defence roles where decision making is important.

11.3.1. Accuracy:

The results from the Experimental Work provided mixed results in relation to personality traits and decision accuracy. Although there were some significant findings related to personality and decision accuracy, in general, the results showed a lack of consistency in the findings in respect of whether these individual differences influenced decision accuracy. The findings support Wallenius et al. (2014) and Saus et al. (2012) who argue that performance is related to more than just personality. Nevertheless, findings from the third experiment (see Chapter 8), showed a positive relationship between accuracy and agreeableness. Individuals who scored more highly on agreeableness had higher levels of decision accuracy, and hence made fewer errors in the task. Individuals who score highly on agreeableness tend to have traits that include being more helpful and mindful (Costa & McCrae, 1992), as well as demonstrating lower levels of distress (Matthew, 1999). Hence, this finding suggests that these individuals were able to apply themselves to the task and deal with the uncertainty of the task without it impacting on their performance. Additionally, Experiment 4 (see Chapter 9) found that decision accuracy was positively related to the personality trait of conscientiousness. Individuals who score higher on the scale of conscientiousness are regarded as being more thoughtful and have a desire to perform well in the task (Costa & McCrae, 1992). Hence, these individuals may have applied themselves more fully to the task, enabling them to make more accurate decisions. Conscientiousness has also been found to be positively related to job performance (Barrick, Mount & Strauss, 1993). Another trait that was associated with performance is neuroticism. In the fourth experiment, higher decision accuracy was related to lower scores of neuroticism. Individuals who score high on neuroticism tend to be less emotionally stable, have a tendency to be anxious and display higher levels of worry. This trait was found to negatively impact on performance in this task. Furthermore, research has shown that cognition is impaired in highly neurotic individuals (Bryne et al., 2015). Hence, the findings demonstrate that high scores on the personality traits of agreeableness and conscientiousness are related to individuals making more accurate decisions, compared to higher scores of neuroticism which were associated with poor accuracy scores. As a result, selecting individuals with these traits may be advantages in air defence roles can make accurate decisions.

Cognitive constructs were also found to relate to decision making. In general, the results related to cognitive constructs varied across the experiments. However, a significant and fairly consistent finding across several of the experiments was that the cognitive construct of Tolerance to Ambiguity was a required trait to assist in making accurate decisions. Those individuals who had a higher Tolerance to Ambiguity scores also made the most accurate decisions. One explanation for this finding is that individuals who are less tolerant to ambiguity are less accurate because they regard the task as threatening. This increases their propensity to give up (Budner, 1962) or it perhaps becomes cognitively overwhelming. The finding that greater tolerance increases the accuracy of the decision also supports the findings of Endres et al. (2009). The implications for this may be helpful when selecting individuals for particular roles, and in particular, in air defence. Consequently, Tolerance to Ambiguity assists people in dealing with uncertain situations and therefore it would be beneficial for operators in decision making roles to be able to tolerate such ambiguities in the environment in which they work. As suggested by Iannello et at. (2017) the ability to deal with uncertainty is beneficial in environments dominated by high WL and ambiguity.

In respect of the populations, although one experiment found NGs to be more accurate than VGP1s, this finding was not replicated in any of the other experiments. Gaming experience did not influence accuracy in the majority of the experiments. As such, further research would need to be conducted to investigate whether playing video games relates to performance in an air defence task, with the possibility of investigating the transfer of skills to other domains. Subsequently, the results into the examination of individual differences in relation to decision accuracy presented in the thesis suggest that characteristics of low neuroticism, high conscientiousness and high agreeableness positively relate to decision making in the air defence task. Thus, the outcomes suggest that individuals who are low on the personality trait of neuroticism and high in conscientiousness may be advantageous. This has led previous authors to argue that these individuals have a resilient personality type (Campbell-Sills et al., 2006). Consequently, a resilient personality type, with an increased Tolerance to Ambiguity may be better able to deal with the uncertainty of air defence decision making. These findings provide support to the idea that certain types of individuals may be better suited to make decisions in uncertain, unpredictable environments, such as air defence.

11.3.2. Confidence:

The results relating to personality and decision confidence were inconclusive, with just one experiment demonstrating a relationship with decision confidence. Experiment 4 demonstrated that neuroticism was negatively related to overall confidence. As discussed earlier, neuroticism may increase feelings of distress and worry and thus reduce confidence in the decision. In the current research, none of the cognitive constructs examined were found to be related to decision confidence. However, there were differences between the groups. The results found that VGPs tended to be significantly more confident in their decisions than NGs across the experiments. However, importantly, this confidence was unrelated to decision accuracy. As such, the reported confidence levels represented overconfidence in decisions. An explanation of these findings can be provided through their experience of playing games, which may have hindered their ability to assess the current situation. Through their experience of game playing, participants may have incorrectly believed they were making accurate decisions. There were also differences between VGP1 and VGP2; those gamers that played more hours were more confident. At present however, there is a dearth of research on the impact of the amount of time spent playing games and how these differences impact on decision making. Interestingly, one previous study by Bailey et al. (2013) found that gamers who play more hours of games have a tendency to make riskier decisions, which could relate to the decision confidence found in this thesis observed. Another explanation for this finding could be that experience, such as playing games, can result in increased decision confidence (Wheatcroft et al., 2017; Atinaja-Faller et al., 2010; Chung & Monroe, 2000). Furthermore, familiarity can also result in increased decision confidence by providing a belief that individuals are accurately remembering important detail (Wheatcroft et al., 2017; Chandler, 1994)

In general, the findings in this thesis demonstrate that confidence in decisions may not be linked to broader personality or cognitive constructs. Decision confidence was however related to playing video games, with VGPs displaying higher levels of confidence in their decisions. This was also more apparent the more hours they spent playing video games. The more hours played, i.e. VGP2s, the more confident individuals were in their decisions. In addition, neuroticism was found to negatively impact decision accuracy. This supports the suggestion that there are types of individuals who may be better suited to make decisions in uncertain, unpredictable environments.

11.3.3. W-S C-A

More recently W-S C-A has been used to examine the suitability of supervisory personnel for UAS pilots (Wheatcroft et al., 2017). Hence, it was envisioned that there was potential for the measure to be applied as a performance measure to more critical environments such as, air defence. In general, the findings from the Experimental Work showed consistent levels of low accuracy scores accompanied by reported high confidence levels. This result could imply that the task was deemed difficult and, as such, an explanation for this could be provided by the Hard-Easy effect (Gigerenzer, Hoffrage & Kleinbolting, 1991). This effect occurs when individuals over estimate their ability by applying higher levels of confidence when the task is difficult. In comparison, when the task is easy, individuals are more likely to attribute lower levels of confidence to their decision. Hence, further work could examine the reduction of the difficulty of the task. As previously discussed, this relationship is also affected by experience and training (Lichtenstein et al., 1977). However, although there was a lack of significance in the findings, the measure has a proven ability in other domains and, as such; future research should consider the use of numerical measures of metacognition.

The research in the current thesis examined whether there were any reported individual differences in relation to metacognition ability. In relation to metacognitive skills, no individual differences were found. This suggests that the measure may not be sensitive enough for these. Contrary to the findings of Wheatcroft et al. (2017), Tolerance of Ambiguity was not found to be related to metacognitive ability as measured by W-S C-A. Further research should examine the individual differences (e.g., personality, cognitive constructs and VGP) that may relate to metacognitive ability.

11.3.4. Groups:

The investigation into the differences between the groups did find some individual differences in personality traits. The analysis on the groups found that VGP2 (gamers who play video games for more than 7 hours a week) were the least conscientious. The trait of conscientiousness is a characteristic of individuals who are more efficient and organised with a desire to do well in the task (Costa & McCrae, 1992). This finding might suggest that this group of VGPs did not engage with the task. In addition, and in support of this finding, the experiments also found that gamers tended to report lower levels of WL. The finding that VGP2s tended to be lower in conscientiousness is interesting, as it supports the existing literature which examines video game addiction. Here it was found that conscientiousness had been negatively related to video game addiction (Wittek et al., 2016). An explanation for this has been provided by Teng (2009), who argues that individuals with low conscientiousness have also been shown to not be able to satisfy needs in the real world; hence, such individuals play video games and may lack the ability to meet the situational demands (Penley & Tomaka, 2001). Alternatively, the fact that VGP2 were overconfident in this task suggests that the gamers may also have been unaware of their lack of ability to meet the demands of the task. However, this study did not control for individuals who may have a gaming addiction. Surprisingly, no differences were found in any other cognitive constructs and could be explained by the different types of cognitive constructs used in previous research. For example, research has tended to look at attentional resources and the attentional visual field (Boot et al., 2008; Hubert-Wallander et al., 2011). Hence, further investigations should consider other measures of cognitive ability. Subsequently, this research has begun to investigate the utility of different populations to investigate critical environments. Crucially, however, the findings contained in this thesis suggest that domain-specific and task-relevant experience is important and which potentially has implications for developing training to improve decision making.

In summary, the current thesis addressed the individual differences in decision making in critical environments, specifically assessing decision accuracy, confidence and metacognitive ability. Overall, the findings into personality traits remain largely inconclusive. However, low neuroticism and high conscientiousness appear to be beneficial in critical environments.

As such, this finding supports previous research that personality may not necessarily be related to performance (Wallenius et al., 2014). Nevertheless, the possibility of a Tolerance of Ambiguity being a key construct in helping individuals make accurate decisions was demonstrated.

11.4. Research Question 3: Does the design, method and measurements used in this thesis align with the wider methods and measurements currently used (SA and WL).

As stated in Chapter three, the current thesis aimed to expand on previous measures of performance and confidence by examining the metacognitive ability of individuals in a critical environment through the application of a novel measure. As such, this project developed an experimental stimulus for the methodological approach which combined objective measures of accuracy, alongside subjective measures of confidence. The aim of the measure was to improve the understanding of air defence decision making and the metacognitive abilities of air defence personnel by combining elements of experimental laboratory testing with NDM methodology. Further, to examine task manipulations against established performance measures in a critical decision making domain, the design was also assessed in relation to WL and SA. It has been argued that these two performance measures are key to military decision making (St John et al., 2000). For the design to be successful, it was important that task manipulations were effective. Hence, WL and SA were used to assess the task manipulations. Furthermore, to increase understanding of performance in air defence decision making, SA and WL were also included in the task to examine their relationship to decision accuracy, confidence and metacognitive ability. These performance measures were also considered in relation to individual differences, the aim of which was to provide a wider view of performance interactions.

11.4.1. Experimental Design:

Performance in critical domains has been assessed in different ways, as discussed in Chapter three; developing an experimental scenario was a key and unique aspect of the current thesis. Akin to *microworlds,* it aimed to bridge the gap between NDM and laboratory settings by allowing the manipulation of variables, development of decision logs and experimental scenarios. Hence, an integrative approach to research was utilized by using SMEs and computer programming techniques to develop a low fidelity scenario for experimental testing.

WL and SA have been found previously to be sensitive to variations in TL and changes in task difficulty are a function of both WL and SA (Selcon et al., 1991). As such, WL and SA were also examined as a means to assess the TL manipulations. The results demonstrate that, in most of the experiments using novice participants, WL was reported as being highest in the high TL condition. These findings suggest that the TL manipulation was successful. Providing support to the findings of Adams-White et al. (2018) and Loft and Sadler (2015), that, higher TL increases feelings of WL. WL, as measured by NASA TLX, also provided an indication as to what subscales of WL influence global WL (i.e., temporal, mental, effort, physical, performance, frustration). By analysing the subsections of WL, participants found that the high TL condition was generally more temporally demanding and required more effort than the low TL condition. With regards to SA, lower scores of SA were reported in the high TL condition. Subscales of SA, in general, revealed that the attentional demand and supply was higher in the high TL conditions with mixed findings with regards to understanding. It could be argued that the finding for high demand and supply in the high TL is an indication of the SART measure being more a measure of WL than SA (Endsley, 1995).

The results in Chapter ten replicated the first experiment with Royal Navy PWO participants. As SA and WL have been demonstrated to be important measures in military decision making, the reported level of SA and WL were also examined in the expert population. Furthermore, it was important that the task demands reasonably matched what would be expected by such an expert. This was to ensure that a suitable level of ecological validity was achieved in the experimental studies. The results from the expert study showed that, in contrast to the novice population, no significant differences in WL or SA were found between the TL conditions. However, by examining the trends in the data, PWOs reported both a higher WL and SA in the high TL condition and WL was far less than that reported by novices. The lower scores of WL in experts suggest that the manipulations would need to be increased to successfully examine the relationship of WL in expert populations. Contrary to novices, PWOs reported having higher SA in the high TL condition. This could be explained by the task being something they are more familiar with, with regards to the speed and information that they were required to attend to. In summary, future work would need to consider further increasing the task demands to replicate real-world decision making when conducting research using experts. This could be achieved by further increasing the task demands for PWO participants. For example, more information that operators need to attend to could be included, as well as additional tracks being presented on the radar screen.

WL and SA were also assessed in relation to time pressure and audio manipulations. The aim was to build an understanding as to how these factors influence performance in air defence decision making. The comparative analysis in Chapter seven found that the main difference between the time to make a decision (10 seconds and 20 seconds) was reported in the SA subscales of attentional supply and demand. That is, reducing the time to make a decision from 20 seconds to 10 seconds increased attention supply and demand of the task. This finding suggests that, although no differences were found on the other scales, SA is sensitive to the time constraints of a task.

In addition, the results showed that audio attendance increased feelings of WL. This finding could be explained by the introduction of an additional audio element increasing the amount of cognitive processing required by the individual and, as such, the experienced WL. This has potentially critical implications for the domain studied. Decision supports have been designed to help reduce the cognitive load on the operator. One way this has been investigated is by providing information via audio. Although previous work has shown that audio can have a negative impact on threat bias (Vachon et al., 2011), the work has not considered the impact on WL on audio air defence decision aids. Hence, these findings suggest that WL should be taken into consideration when implementing audio into decision making aids. However, the experience of having to attend to audio was unrelated to decision accuracy, confidence and W-S C-A and did not have an adverse impact on decision making. In light of the comments made, further research into the effects of audio presentation is warranted to consider the full implications on air defence decision making

As part of the current thesis, the research conducted in Chapter nine introduced a metacognitive feedback training element. As previously stated, this thesis aimed to examine performance measures in a critical domain. As SA and WL are important performance measures, the relationship between these performance measures and metacognitive feedback training was investigated. Previous research has demonstrated that SA can be improved via critical thinking and training skills, which help metacognitive ability (Cohen et al., 1998). However, the findings in this research found that MFT had no effect on SA. However, the MFT applied specifically related to individual's awareness in the accuracy of their response, not tailored towards improving SA. Hence, future work could examine the types of training which would be needed, for instance, improving critical skills and/or domain specific training which could target improved SA.

WL was also explored, the results demonstrating that individuals who received MFT did not report higher levels of WL. Thus, although W-S C-A discrimination was higher in these individuals, this did not impact on their feelings of WL. This finding implies a positive effect of MFT, as it increased an individual's ability to discriminate between accurate and inaccurate responses without overloading the individual. This supports the previous findings from the other experiments in the thesis, as W-S C-A was found to be unrelated to WL. An explanation could, therefore, be that WL and W-S C-A rely on different cognitive resources. As discussed in Chapter nine, WL is related to an individual’s attentional resources and is subjective to the individual. On the other hand, W-S C-A is based both on subjective and objective metrics. This thesis has begun to examine some of the factors involved in this, and provide insight into the cognitive resources and the factors that mediate the relationship between confidence and accuracy.

11.4.2. W-S C-A - Measurement:

It was envisioned that the measure could be applied as a performance measure to more critical environments such as, air defence. However, W-S C-A was not found to relate to WL or SA. As such, based on the results reported in this thesis, further investigation into metacognition using this measure alongside WL and SA may be necessary. This is contrary to the findings of Kim et al. (2018), who found a negative relationship between metacognition and WL. Although WL and SA were found not to be related to W-S C-A, the results did demonstrate these performance measures to be related to decision confidence and accuracy, in addition to some individual differences.

11.4.3. Accuracy and Confidence:

WL and SA were also assessed in relation to decision accuracy and confidence. A consistent finding was that WL was found to be related to a decrease in decision confidence. This is an important finding for decision making, as reduced confidence in decisions taken could lead to increased WL. Individuals may thus seek out more information to support/contradict decision certainty. Crucially, however, WL was unrelated to decision accuracy thus, suggesting that WL is subjective and not related to objective accuracy. In comparison, SA was related to an increase in decision confidence. This supports the findings of Adams-White et al. (2018). Here individuals who reported higher levels of SA were also more confident in their decisions. However, these findings should be regarded with caution. SA was measured subjectively and a confidence bias has previously been found in SA reporting (Sulistyawati & Chui, 2009). Thus, it might be that individuals are generally confident in their assessments of SA performance.

Importantly, SA was not related to accuracy in decisions taken. Therefore, the confidence displayed by the individuals may indicate that individuals privately believe that they had a better understanding of the situation than they accepted. This supports Endsley's (1995) argument that self-ratings of SA tend to be related to a statement of how certain the person feels about SA. Unlike, Stanners and French (2005), SA was found not to be related to decision accuracy. The lack of significance in SA could also be due to the level of experience of the participants in the task. Further, it has been argued that SA is a domain dependent cognitive construct (O'Brien & O'Hare, 2007). It is therefore important when investigating performance measure such as WL and SA in air defence decision making that decision confidence is considered.

11.4.4. Individual Differences in WL and SA:

The outcomes demonstrate that there may be individual factors that can influence people's feelings of WL and SA. This can provide valuable insight into selecting an individual for certain roles and tailoring training to suit individual needs. High SA and low WL may have a positive impact on decision making as it allows individuals to be able to assess the situation without limiting their cognitive resources. One individual difference which may hold prominence is the influence of VGP (Wheatcroft et al., 2017). The research in the thesis showed mixed findings in relation to increased SA and reduced WL in the gaming populations. There was some evidence to suggest a positive impact i.e. higher SA and lower WL. Indeed, VGPs tended to report lower levels of WL experience during the task. This could imply that they are better suited to the demands of the task. However, this was not a consistent finding across all the experiments.

Differences in SA were also reported. In Experiment 3, an interaction was found between group and audio in reports of SA. Outcomes showed that VGP2s reported higher levels of SA than VGP1 and NGs when they were instructed to attend to the audio. It could be that attending to the audio increased VGP2s belief in their SA. It has been previously demonstrated that gaming and the presence of audio is beneficial to feelings of SA (Chiappe et al., 2013). Furthermore, research has also shown that game play has been linked to the widening of attention whilst completing the task (Green & Bavelier, 2003), which could have also influenced SA in game players. In comparison, NGs reported similar SA levels when attending and not attending to the audio.

However, on the contrary, Experiment 1 found the absence of a relationship between gaming experience and SA. This finding provides further support to the findings of Vidulich et al. (1995), who found that game experience did not improve SA. Again, an explanation could be that SA is domain-specific and hence task-relevant experience is important (O'Brien & O'Hare, 2007). Although video game experience is beneficial to some aspects of cognitive ability, SA and WL, it would be beneficial for further research to assess the suitability of gamers as a research population. This would allow for the investigation of the impact of task relevant training and the transferability of skills to other domains. Future work could also consider the direct relationship with the subscales of WL and SA (Chiorri et al., 2015). For instance, Experiment 2 found that in the subscales of WL, NGs reported more mental demand. Thus, there is evidence to suggest that, in relation to WL, certain individual traits might be suited to certain job roles in the domain studied.

The research also found relationships between observed personality traits and WL and SA. Although not universal in the findings, SA was found to be related to conscientiousness and extraversion in experiment 2. As discussed, extraversion and self-reported SA have been linked to confidence. In Experiment 3, WL was found to be related to agreeableness and conscientiousness. These findings imply that there are positive traits that may assist individuals in dealing with demanding environments such as air defence. The findings from the thesis would suggest that understanding individual differences in WL and SA warrants further investigation.

11.5. Research Question 4: Can Metacognitive ability in air defence be improved through feedback training?

Section B of the Experimental Work assessed whether metacognitive ability could be improved via a feedback training (MFT) session. Thus, first, as part of the research, it was important to understand and apply the research findings to explore how the measure might be relevant to the development of a potential Toolkit in the domain. Second, with regards to the application, it was necessary that the findings from the novice population were generalisable to experts. The previous experiments demonstrated overconfidence in decisions with a poor ability to discriminate between accurate and inaccurate decisions. This was indicated by low W-S C-A scores with a tendency for overconfidence, as demonstrated by percentage confidence analysis. Previous research has found that overconfidence can be improved via different training methods (Hoffman & Spatariu, 2008; Fiore et al., 2010; Kim, 2018). Indeed, the results and discussion of Chapter nine demonstrated that the introduction of a small training and feedback session at the start of the experimental task increased confidence in correct decisions. Although, confidence remained high, the improvement in metacognitive ability, as assessed by W-S C-A is a key finding. However, as the MFT was conducted on novice participants, this finding could be explained by the inability to apply metacognitive strategies to the task due to the air defence scenario being unfamiliar to them (Barnett & Ceci, 2002). Further work could be carried out to explore this finding with the use of a greater number of experts in this field. The implications of training are further discussed in Chapter twelve. In addition, the finding that MFT increased W-S C-A is useful and important finding in building upon the understanding of decision making in air defence and in developing a toolkit.

Another key aspect to investigate in the thesis was to increase the understanding of how experts make decisions and the application and generalisability of the research findings. In accordance with the NDM criteria set forward by Johnston et al. (1998), Chapter ten conducted research on recently trained PWOs with on average 8 years of experience in the RN. The results provided support for the findings demonstrated in the novice populations in the previous experiments. The results also showed that DC was found to impact on decision accuracy and confidence. PWOs made more accurate decisions when presented with medium DC events compared to the low DC events and decision confidence was higher in those low DC events. This finding provides further evidence that DC is a crucial aspect to decision making in air defence and warrants further investigation. However, no other findings were found. TL did not impact on decision accuracy or confidence and no differences were found in feelings of WL or SA in the PWO populations. An explanation for this could be due to the small sample size collected. As previously discussed, differences in WL and SA were not necessarily sensitive to the task manipulations. As such, caution should be applied to the findings.

11.6. Summary

The key findings from the Experimental Work are highlighted in the table below:

Table 52: Summary of key findings

Decision Criticality	• Decision Criticality is one of the main contributors to performance. High Decision Criticality was found to positively impact on performance by increasing decision accuracy. • Decision Confidence was most impaired by a Medium Criticality. • Overall Decision Criticality is key aspect of decision accuracy and confidence which warrants further investigation in critical decision making environments.
Individual differences	• Video Game Players consistently displayed higher confidence in their decisions. However, this did not necessarily map on to the accuracy of their decisions. • Individuals who had a higher Tolerance to Ambiguity were more accurate in their

	decisions highlighting that being more tolerant may be beneficial in dealing with uncertain environments.
Application of MFT	• The introduction of a small scale metacognitive feedback element was beneficial in improving metacognition. Further research should examine the use of metacognitive training in decisions made in critical environments.

CHAPTER TWELVE

12. Limitations, Implications, Recommendations and Conclusions

No research is without its limitations. This section discusses the limitations of the research presented, as well as the implications, recommendations and future work based on the current findings. The methodological approach in the current thesis integrated both NDM and CDM theories. Drawing on a range of current thinking to inform the methodological stance has demonstrated that an integrative approach is useful to the decision making research domain and provides valuable insights to the benefit of both approaches. This Chapter discusses the implications for research which might include, for example, assistance for decision support that could highlight and identify the dangers of overconfidence in certain situations. The current work contained in this thesis points toward further research into MFT to improve trainee and operator metacognitive insight to ensure the correct level of confidence is applied to decisions. The Chapter also contains researcher reflections. First, the potential limitations of the work will be discussed.

12.1. Limitations

12.1.1. Participants: As discussed in the discussion of Chapter five, most of the Experimental Work was conducted using novice participants. Novice participants had no prior experience of air defence decision making. As stated in Chapter two, the method in the current thesis has integrated the CDM and NDM approaches. In doing so, it was important to conduct research on larger scale populations than could be obtained by using experts alone. Although expert participants provide higher ecological validity, the availability of such experts to conduct large scale experimental work in this manner is extremely limited. Hence, the generalisation of the findings in the Experimental Work conducted using novices may be limited and an awareness of the potential differences between expert and novices when conducting and applying the research to experts should be acknowledged. Nevertheless, there are benefits of conducting research using novice participants. For instance, it allows for a better understanding on the development of expertise. Indeed, a similar method has recently been developed by the ShadowBox method (Klein et al., 2013) discussed in Chapter three. Nevertheless, the thesis aimed to address this limitation in Chapter ten by conducting research using experts, to assess how decision making aligns with findings from the novice population. Indeed, the results of this experiment conducted on PWOs (experts) demonstrated similarities between the findings of the experts and novices. DC impacted on confidence and accuracy in a similar way. Medium and high criticality decision events were made with more accuracy and low DC events made with more confidence. Although, no other similarities were found, the expert study was conducted on a small sample and future work should consider a larger sample size as well as a wider range of experience levels.

The work contained in the current thesis sought to integrate approaches from NDM and CDM. Hence, in the experiments, variables were manipulated (e.g., the time to make a decision) which allowed for greater control of the experiment and the isolation of factors, however, it can reduce the ecological validity of the research. Even so, the benefits of doing so have provided new insights into factors such as DC. In addition, participants were presented with option choices resulting in individuals being forced into taking up more analytical decision making strategies. In keeping with the natural time pressures and procedures used in air defence, it was believed that these factors were beneficial to increase internal validity. Nevertheless, this research may also benefit from the use of more qualitative measures, such as, interviewing, to gauge a wider view of the decision making process involved in decision criticality.

The experimental scenario in the current thesis was developed in collaboration with SMEs to provide an in-depth, "gold standard" response, as well as a novel input into the design of experimental work. This approach had been previously validated in the air pilot domain (Wheatcroft et al., 2017) as well as in the air defence domain (Adams-White et al., 2018). It was therefore considered to be a useful way forward. As discussed in Chapter three, similar methods including CTA can then be used to design simulations and *microworlds;* thus, supporting the benefits of experts working with researchers to develop unique insights and new methodologies. The original experiment was designed to be tested on experts. Consequently, the design was not developed with the layperson in mind. Throughout the experiments, the fact that low accuracy scores were found in both novice and expert decision makers, and which was also accompanied by elevated decision confidence, would suggest that the scenario task was difficult for both novice and experts. In its ambitious nature, this could have created some difficulty in the complexity of the task. Nevertheless, steps were taken to provide individuals with the relevant task information to make their decisions. These steps included providing all participants with the same briefing documents. These provided the individuals with the knowledge necessary to assist them with the decision making in the experimental tasks.

Furthermore, one possible explanation for the expert findings being relatively similar to that of novices could be that the scenario was novel to both groups of participants. As such, the expected expertise of the PWOs was not accessible to them during the experiment. The low accuracy scores recorded for the PWOs could also have arisen from them trying to "fight" the scenario and fidelity. As such, they did not apply themselves to the task. This could have occurred if the experimental scenario did not behave in a manner in which they expected. Nevertheless, these findings would support those of Chapman et al. (2006) who found no differences between experts and non-experts. This raises the question of the impact of fidelity on experimental research. In contrast, research has demonstrated that high fidelity may not always be necessary (Walker, Takayama & Landay, 2002). Further research could examine the impact of fidelity on performance in the task. This will be discussed later in this Chapter.

12.1.2. Measures and Materials: W-S C-A is a quantitative measure of metacognition, which is a specific calculation of the relationship between subjective confidence and objective accuracy. According to Fleming and Lau (2014), the relationship between decision confidence and accuracy can provide a quantitative measure of metacognition; thus, this thesis examined metacognition quantitatively. Although previous measures have been used to examine metacognition, this measure provided a unique insight to metacognition in the air defence domain and its relationship to other factors relevant to decision making. An interesting finding was that mean W-S C-A scores remained fairly low across experiments. Further research into how this measure could be developed is needed to be able to successfully use this as a measure of metacognition in this domain. As demonstrated in Chapter nine, metacognitive ability can be improved by the application of a short training session (MFT). Future work could examine a more in depth training session. Recent research is currently investigating this through adaptive training skills (Ward, Gore, Hutton, Conway & Hoffman, 2018 - see section 12.3.2).

As might be expected, the limitations of this research also include the self-reported nature of personality, WL and SA, which subsequently might have been influenced by respondent bias and/or self-presentation style (Spector & O'Connell, 1994). Consequently, the use of more objective measures would be beneficial, for instance the Situation Awareness Global Assessment Technique (SAGAT), which is a freeze online probe technique aimed at measuring SA (Endsley, 1988). Using this method, at certain time intervals, the scenario would be paused and an individual would be asked questions relating to their perception of the situation. However, this requirement relies on expensive simulations and the analysis requires extensive preparation (Stanton, Salmon, Walker, Baber & Jenkins, 2005). In addition, physiological measures could be used to examine TL to provide more objective measurements. Despite this, it remains that subjective measures provide insight into influencers known to impact on performance and decision making.

12.2. Implications

12.2.1. Theory: In Markman's (2018) paper, the author calls for an integrative approach to decision making research by combining approaches of CDM and NDM. As pointed out by Lipshitz et al. (2001), mixed methods could be used to help make predictions and build better theoretical understanding. Moreover, there is a current interest in the improvement of the credibility and transferability of NDM research (McAndrew & Gore; 2013; Roberts & Cole, 2018), which may be sought through the introduction of more experimental work. In line with this movement, the thesis aimed to bridge the gap between NDM and a more experimental measure. The research in this thesis used a range of subjective and objective measures to try to enhance the understanding of decision accuracy, confidence and metacognitive ability. The scope was to increase understanding of decision making by examining the relationships between the environment and the individual using a mixed methodology. Future research, in critical environments could assess these dynamics using other measures. For example, recent research has investigated this mixed method approach in Authorised Firearms Officers (AFOs; Roberts & Cole, 2018).

12.2.2. Population: The thesis also provided implications on the types of research participants and the potential benefits of using different populations to conduct research. As mentioned in Chapter two, VGPs have different skill sets and have been found to demonstrate skills which would be relevant to the critical environment of air defence. Thus, the use of VGPs provides new insights into the use of this population in NDM research. Indeed, these populations could be used to investigate how these perceptual skills are mapped onto expertise (Elliott et al., 2007). In addition, the research in the thesis demonstrated that, in general, gamers tended to have lower WL and higher SA which could suggest their suitability for conducting more research on gamers as they map onto similar traits found in experts. However, as the findings suggest, task relevant experience is an important consideration when investigating different populations.

12.2.3. Decision Support and Automation: Decision making in air defence is complex and demanding and is based on a range of different sensory modalities. Decision supports are generally designed to assist individuals with decision making tasks. With regards to air defence decision making, research has considered the reduction of WL by assisting operators through audio support, i.e. detect changes in criticality via audio (Vachon et al., 2011). The results from Chapter eight may also have implications on the design of decision support systems as they demonstrated that attending to an audio cue increased individuals experience of WL. Hence, when designing decision supports, it is important that the individual's subjective feeling of WL is considered. Furthermore, using the findings from this research, it would be beneficial for decision supports to highlight and identify the dangers of overconfidence in certain situations i.e. in low criticality decisions. In addition, more and more decisions in an Ops room are being automated to help ease the demands on the operator. Although there is an abundance of research on the impact of trust in automation, very little is known about the impact automation has on an individual's metacognitive ability. Wheatcroft et al. (2017) found that individuals tended to apply higher levels of confidence in decisions which were automated, compared to their decision to manually control a UAS. This has implications on future actions, as the level of confidence in automated decisions was not necessarily linked to the accuracy of the decision. Although one focus of the current thesis was to assess the decision making of PWOs, increasingly, decisions are being made by automated systems. Hence, the method used here could be used to examine metacognitive ability in relation to the use of automation, as well as decision support tools.

12.2.4. Team Metacognition: Individual metacognition is an important aspect of decision making. However, metacognition can also be applied to the level of a team. Decision making in air defence is multi-faceted and occurs on multiple levels from the individual, to the team, to the environment; hence, it is important to understand decision making at multiple levels. It has been argued that, in an Ops room, there is also a need for an understanding of the group process with regards to cognition and metacognition. Thus, although the PWO is the main decision maker and the expert focus in the current thesis, many of the previous decisions which lead up to their decision events will be conducted by other members of the team. Hence, a PWO's decision is based on concepts such as shared SA and metacognition (Salas & Fiore, 2004). It would therefore be beneficial to consider the metacognitive ability of groups in an air defence environment. Indeed, research has shown that the confidence of another person can impact on an individual's confidence (Sniezek, & Henry, 1989). Displaced confidence has implications on future actions when not related to the accuracy of those decisions.

12.2.5. Fidelity: The research conducted in this thesis used a low physical fidelity computer simulated experiment based on information provided by SMEs. As mentioned in Chapter three, the aim was to examine the psychological fidelity and therefore assess psychological constructs and cognitive mechanisms. Although attempts to increase the physical fidelity of the work were assessed in experiments two and three, future work could consider the impact of physical fidelity in the task in more detail. This may involve working with pre-existing training in the available training centres and via the use of greater physical fidelity simulations.

12.3. Recommendations

12.3.1. Toolkit: As suggested in Chapter one, the research contained in this thesis sought to take the first steps toward the development of a Toolkit to measure decision accuracy, confidence and metacognitive abilities of air defence operators. As such, at this stage, three different computer generated training scenarios have been developed and created which range in TL and DC. In addition, these are accompanied by a set of agreed decision logs with SMEs, agreed decision events, and the corresponding accurate decisions. It is intended that the future development of the work would use these to assess the levels of confidence, accuracy and W-S C-A relationships for current and training Ops room decision makers. For this to occur, the decision and related confidence responses of individuals to the scenarios presented need to be recorded. The individual decision responses are assessed against the previously agreed decision logs. Once enough data is collected, the decisions can then be categorised. The researchers can work with those interested to develop the appropriate categorical data required. Future work would need to continue to examine psychological traits that are the most beneficial to the effective performance of personnel and which may relate to higher levels of accuracy together with relationships between confidence and accuracy.

12.3.2. Training: The tools for examining associated confidence in a decision could easily be applied to current training techniques, which would enable the examination of metacognitive awareness**.** As mentioned, the approach has been successfully used in the suitability of UAS pilots (Wheatcroft et al., 2017) and to examine the impact of questioning types in forensic settings (Wheatcroft & Woods, 2010). Until now, the application to training using this measure in air defence has not been researched. It is apparent that individuals are largely unaware of the impact confidence may have on decisions and actions taken. Hence, greater awareness is needed into confidence and overconfidence in the decision taken in the air defence domain.

As demonstrated in Chapter ten, providing a small training session and relevant individual feedback on an individual's metacognition showed a significant and positive impact on their metacognitive awareness. That being said, metacognitive scores still remained relatively low. As demonstrated in previous research, more in-depth training and feedback can increase metacognitive awareness (Cohen et al., 1995; Kim, 2018). Furthermore, it has been argued that using metacognition in training can be used to develop a deeper understanding of the causal factors to effective decision making (Ward et al., 2018). As such, the findings of this thesis provided evidence that metacognition should be included in any future training protocols in the air defence context to help trainees apply the appropriate levels of confidence to both accurate and less accurate decisions. Of course, future work would need to apply any training to experts. Again, it has been argued that the ability to be metacognitively aware is beneficial to the development of expertise, as expertise requires continual development and learning, which, in turn, relates to metacognitive awareness (Ward et al., 2018; Fadde & Klein, 2010). Hence, future work would need to consider the development of a model of metacognitive awareness training.

Additionally, future training needs will need to take into consideration the impact of different types of decisions; that is, criticality. For example, the work contained in this thesis could be used to inform individuals of when they may make a highly confident yet incorrect decision (i.e., in low DC). As well as making them aware of the likelihood of making errors in less critical decisions. The necessary visual systems could be developed to alert operators to these conditions and to assist with real-time feedback.

12.4. Researcher Reflections on Experimental Research in the NDM Domain

There is no doubt that conducting research with experts in field settings is highly beneficial and desirable. However, there are disadvantages to this type of research, as discussed in the review of Chapter three. For instance, some of the methods require training, extensive reliance on SMEs, or are time-consuming to conduct. Although high fidelity *microworlds* bridge a gap and allow for more experimental testing using experts, they also require specialist computer programming knowledge or access to pre-existing programmes. Thus, conducting research to align to the criteria of NDM remains a difficult task for researchers with limited availability and access to experts. One way this thesis addressed this issue was by the development of an experimental set up which could be conducted by both novices and experts. However, as discussed earlier, generating a task that is applicable to both populations is complex and thus the applicability of the findings to both populations may be reduced. That said, in order to move forward with NDM research (which aims to build an understanding and further develop the field), it is essential that greater access to domain expertise is generated by practitioners to facilitate researchers in the opportunity to conduct more experimental work in the NDM paradigm, together with greater availability of training in NDM methodological approaches (e.g., CTA). Thus, practitioners assisting researchers at all levels should be positively encouraged and could also include working alongside research conducted "in house" to explore potential avenues to increase accessibility to expertise in the relevant domain.

Overall, as previously stated by some authors, the aim should be to improve the credibility and transferability of methods used in NDM

(McAndrews & Gore, 2013; Roberts & Cole, 2018) and this is one way in which it could be facilitated. Importantly, it must be highlighted that the methods proposed in the thesis and use of a range of different methods should be used in conjunction with NDM methods, not in opposition

12.5. Conclusions

Air defence decision making is characterised by time pressure, complexity and uncertainty. It is important that accurate and appropriately confident decisions are made. To conclude, the fundamental problem that this thesis addressed was to increase the understanding and measure of decision confidence and accuracy in air defence decision making. To do so, a novel method was designed and developed which used an integrative research methodological approach which was based on CDM and NDM theories of decision making. The thesis has made a unique contribution to research into the decision making and metacognition literature by increasing understanding of factors that influence decision making in the air defence domain.

First, the research includes the first study to use this novel method to assess decision making in RN air defence personnel, with the results highlighting some similarities across novice and expert participants.

Second, the findings from the research contain herein provide valuable insight into the external and internal factors that relate to air defence decision making. Importantly, the work clearly showed DC as an important factor that needs to be addressed when investigating decision making in critical environments. The approaches and techniques used here may also be beneficial to other similar critical environments such as other command and control settings, aviation and healthcare.

Third the thesis has expanded on performance measures by investigating the relationship between decision accuracy and confidence and the relationship between the two (i.e., metacognition). Thus, demonstrating how these metrics can be used to guide decision making and the benefits of increasing operator's awareness of their metacognition to improve learning and incorporate into training and development. Indeed, Chapter nine demonstrated that operator's awareness of their metacognitive ability (W-S C-A) can be increased using a simple training event (MFT). In addition, there has been a focus on the individual decision maker and specifically, metacognitive ability. By adopting an individual perspective, the thesis has added merit to investigating individual differences in air defence operators. The findings suggest that there are certain personality traits and cognitive constructs that influence decision making and may be of benefit for job role selection.

Finally, the outcomes of this thesis provide the building blocks for the development of a Toolkit for air defence operator's decision accuracy, confidence and metacognition. Taking into account the methodological approach which combines experimental paradigms with NDM research, the premise upon which the Toolkit is built may also be applied to wider critical environments. It seems certain that firmly forging these links in future investigations will be of benefit to the organisations which seek to successfully employ the outcomes.

References

Adams-White, J. E., Wheatcroft, J. M., & Jump, M. (2018). Measuring Decision Accuracy and Confidence of Mock Air Defence Operators. *Journal of Applied Research in Memory and Cognition*, *7*(1), 60-69.

Adams-White, J. E., Wheatcroft, J. M., & Jump, M (2017). Measuring the meta and cognitive abilities of air defence operators. In: 13th International Naturalistic Decision Making Conference, 2017-06-20 - 2017-06-23, Bath.

Allen, J. A., Hays, R. T., & Buffardi, L. C. (1986). Maintenance training simulator fidelity and individual differences in transfer of training. *Human Factors, 28*(5), 497-509.

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410.

Atinaja-Faller, J., Quigley, L., Banichoo, S., Tsveybel, V., & Quigley, M. (2010, November). Simulation training effect on confidence levels of senior clinical decision-making nursing students in managing an acute event. Paper presented to International Conference of Education, Research and Innovation, Madrid.

Bailey, K., West, R., & Kuffel, J. (2013). What would my avatar do? Gaming, pathology, and risky decision making. *Frontiers in psychology*, *4*, 609.

Banbury, S., Fricker, L., Tremblay, S., & Emery, L. (2003). Using auditory streaming to reduce disruption to serial memory by extraneous auditory warnings. *Journal of Experimental Psychology: Applied*, *9*(1), 12.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist*, *28*(2), 117-148.

Barnes-Yallowley, J. (2015). Subject matter expert personal communication. Somerset: SEA Ltd.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612.

Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of applied psychology*, *78*(5), 715.

Bartels, D. M., Hastie, R., & Urminsky, O. (2018). Connecting Laboratory and Field Research in Judgment and Decision Making: Causality and the Breadth of External Validity. *Journal of applied research in memory and cognition*, *7*(1), 11-15.

Beach, L. R., & Lipshitz, R. (2017). Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. *Decision making in aviation*, *85*, 835-847.

Beaman, C. P., Hanczakowski, M., & Jones, D. M. (2014). The effects of distraction on metacognition and metacognition on distraction: evidence from recognition memory. *Frontiers in Psychology*, *5*, 439.

Bliss, J. P., & McAbee, P. E. (1995). Alarm responses in a dual-task paradigm as a function of primary task criticality. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 39, No. 21, pp. 1395-1399). Sage CA: Los Angeles, CA: SAGE Publications.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta psychologica*, *129*(3), 387-398.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2-3), 171-184.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*(1), 11.

Budner, S. (1961). An Investigation of Intolerance to Ambiguity. *Nursing Research*, *10(3)*, 179.

Buratti, S., Allwood, C. M., & Kleitman, S. (2013). First- and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition and Learning*, *8(1)*, 79–102.

Burns, K. M., Burns, N. R., & Ward, L. (2016). Confidence—More a personality or ability trait? It depends on how it is measured: A comparison of young and older adults. *Frontiers in psychology*, *7*, 518.

Byrne, K. A., Silasi-Mansat, C. D., & Worthy, D. A. (2015). Who chokes under pressure? The Big Five personality traits and decision-making under pressure. *Personality and individual differences*, *74*, 22-28.

Calfee, S. H., & Rowe, N. C. (2004). Multi-agent simulation of human behavior in naval air defense. *Naval Engineers Journal*, *116(4),* 53-64.

Callister, J. D., Percival, G. L., & Retzlaff, P. D. (1999). Stress and fatigue during aircrew survival training. In *International Symposium on Aviation Psychology, 10 th, Columbus, OH* (pp. 843-847).

Campbell-Sills, L., Cohan, S. L., & Stein, M. B. (2006). Relationship of resilience to personality, coping, and psychiatric symptoms in young adults. *Behaviour research and therapy*, *44*(4), 585-599.

Cannon-Bowers, J. A., & Salas, E. E. (1998). *Making decisions under stress: Implications for individual and team training*. Washington DC: American Psychological Association.

Chamberland, C., Hodgetts, H. M., Vallières, B. R., Vachon, F., & Tremblay, S. (2018). The Benefits and the Costs of Using Auditory Warning Messages in Dynamic Decision-Making Settings. *Journal of Cognitive Engineering and Decision Making*, *12*(2), 112-130.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & cognition*, *22*(3), 273-280.

Chandra, S., Sharma, G., Salam, A. A., Jha, D., & Mittal, A. P. (2016). Playing Action Video Games a Key to Cognitive Enhancement. *Procedia Computer Science*, *84*, 115-122.

Chapman, T., Nettelbeck, T., Welsh, M., & Mills, V. (2006). Investigating the construct validity associated with microworld research: A comparison of performance under different management structures across expert and non-expert naturalistic decision-making groups. *Australian Journal of Psychology, 58*(1), 40-47.

Chiappe, D., Conger, M., Liao, J., Caldwell, J. L., & Vu, K. P. L. (2013). Improving multi-tasking ability through action videogames. *Applied ergonomics*, *44*(2), 278-284

Chiorri, C., Garbarino, S., Bracco, F., & Magnavita, N. (2015). Personality traits moderate the effect of workload sources on perceived workload in flying column police officers. *Frontiers in psychology*, *6*, 1835.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*(1), 83-117.

Chung, J., & Monroe, G. (2000). The effects of experience and task difficulty on accuracy and confidence assessments of auditors. *Accounting & Finance*, *40(2)*, 135-151.

Clarke, *F. R.*, Birdsall, *T. G.*, *&* Tanner, *W. P. Jr. (*1959*). Two types of ROC curves and definitions of parameters. Journal of the Acoustical Society of America*, *31*, 629-630.

Cohen, M, S. (1980). Aftereffects of stress on human performance and social behavior: a review of research and theory. *Psychological bulletin*, *88*(1), 82.

Cohen, M. S., Freeman, J. T., & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. *Making decisions under stress: Implications for individual and team training*, 155-189.

Colbert, A. E., Barrick, M. R., & Bradley, B. H. (2014). Personality and leadership composition in top management teams: Implications for organizational effectiveness. *Personnel Psychology*, *67*(2), 351-387.

Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and anxiety*, *18*(2), 76-82.

Corso, K. B., & Löbler, M. L. (2011). Understanding the subject's behavior in the interaction with a decision support system under time pressure and missing information. *JISTEM-Journal of Information Systems and Technology Management*, *8*(3), 581-604.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Craig, D., Morales, D., & Oliver, M. (2004). *USS Vincennes incident* [PDF]. Retrieved from https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-422-human-supervisory-control-of-automated-systems-spring-2004/projects/vincennes.pdf.

Crandall, B., Klein, G., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: MIT Press.

Cumming, S. R., & Harris, L. M. (2001). The impact of anxiety on the accuracy of diagnostic decision-making. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *17*(5), 281-286.

Dahlstrom, N., Dekker, S., Van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science*, *10*(4), 305-314.

Darke, S. (1988). Anxiety and working memory capacity. *Cognition and emotion*, *2*(2), 145-154.

Darr, W. (2011). Military personality research: A meta-analysis of the Self Description Inventory. *Military Psychology*, *23*(3), 272-296.

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, *78*, 97-119.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship?. *Law and Human Behavior*, *4*(4), 243.

Dekker, S. W., Hummerdal, D. H., & Smith, K. (2010). Situation awareness: some remaining questions. *Theoretical Issues in Ergonomics Science*, *11*(1-2), 131-135.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological science*, *29*(5), 761-778.

DiBonaventura, M. D., & Chapman, G. B. (2008). Do decision biases predict bad decisions? Omission bias, naturalness bias, and influenza vaccination. *Medical Decision Making*, *28*(4), 532-539.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, *41*(1), 417-440.

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Efklides, A. (2001). Metacognitive experiences in problem solving. *Trends and Prospects in Motivation Research*, *2*, 297-323.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological bulletin*, *73*(3), 221.

Elliott, T., hew Welsh, M., & Mills, V. (2007). Investigating naturalistic decision making in a simulated microworld: What questions should we ask? *Behavior Research Methods, 39(4)*, 901-910.

Endres, M. L., Chowdhury, S., & Milner, M. (2009). Ambiguity tolerance and accurate assessment of self-efficacy in a complex decision task. *Journal of Management & Organization*, *15*(1), 31-46.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1)*, 32-64.

Endsley, M., & Jones, W. M. (1997). *Situation Awareness Information Dominance & Information Warfare*. In *Technical Report 97-01*, Belmont, MA: Endsley Consulting.

Erjavec, J., Khan, N. Z., & Trkman, P. (2016). The Impact of Personality Traits and Domain Knowledge on Decision Making-a Behavioral Experiment. In *European Conference on Information Systems*. Research-in-progress papers.38.

Eubanks, J. L., & Killeen, P. R. (1983). An application of signal detection theory to air combat training. *Human factors*, *25*(4), 449-456.Fadde, P. J., & Klein, G. A. (2010). Deliberate performance: Accelerating expertise in natural settings. *Performance Improvement*, *49*(9), 5-14.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39(2)*, 175-191.

Finley, J. R., Benjamin, A. S., & McCarley, J. S. (2014). Metacognition of multitasking: How well do we predict the costs of divided attention?. *Journal of experimental psychology: applied*, *20*(2), 158.

Fiore, S., & Vogel-Walcutt, J. J. (2010, September). Making metacognition explicit: Developing a theoretical foundation for metacognitive prompting during scenario-based training. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 27, pp. 2233-2237). Sage CA: Los Angeles, CA: SAGE Publications.

Fiorella, L., Vogel-Walcutt, J. J., & Fiore, S. (2012). Differential impact of two types of metacognitive prompting provided during simulation-based training. *Computers in Human Behavior*, *28*(2), 696-702.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34(10)*, 906.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541-1543.

Flin, R. (2001). Selecting the Right Stuff: Personality and high-reliability occupations. In: Roberts, R.W. and Hogan, R. (eds), *Personality Psychology in the Workplace*. Washington, DC: American Psychological Association, pp. 253–276.

Fogarty, W. M. (1988). Formal investigation into the circumstances surrounding the downing of a commercial airliner by the U. S. S.

Vincennes (CG 49) on 3 July 1988 [Unclassified Letter Ser. 1320 of 28 July 1988, to Commander in Chief, U.S. Central Command]. Washington, DC: U.S. Department of the Navy.

Friedrich, M., Biermann, M., Gontar, P., Biella, M., & Bengler, K. (2018). The influence of task load on situation awareness and control strategy in the ATC tower environment. *Cognition, Technology & Work*, *20*(2), 205-217.

Frye, L. M., & Wearing, A. J. (2014). What were they thinking? A model of metacognition for bushfire fighters. *Human Factors Challenges in Emergency Management: Enhancing Individual and Team Performance in Fire and Emergency Services*, *57*.

Ghosh, D., & Ray, M. R. (1997). Risk, ambiguity, and decision choice: Some additional evidence. *Decision Sciences*, *28*(1), 81-104.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart* (pp. 3-34). Oxford University Press.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological review*, *98*(4), 506.

Gonsalves, J, D. (1997). *The Tactical Decision Game (TDG): An Invaluable Training Tool for Developing Junior Leaders TDGs* Retrieved March 2015 from http://ciar.org/ttk/mbt/armor/armor-magazine/armor-mag.1997.mj/3tdg97.pdf.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, *46*(3), 449-460.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in human behavior*, *21*(2), 273-286.

Gore, J., Flin, R., Stanton, N., & Wong, B. W. (2015). Applications for naturalistic decision-making. *Journal of Occupational and Organizational Psychology*, *88*(2), 223-230.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, *2*(2), 205-227.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology*, *24*(3), 411-435.

Hammer, C. H., & Ringel, S. (1965). The Effects of Amount of Information Provided and Feedback of Results on Decision Making Efficiency1. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *7*(6), 513-520.

Hammond, K. R. (2000). *Judgments under stress*. Oxford University Press.

Hanson, J. A., Bliss, J. P., Harden, J. W., & Papelis, Y. (2014). The effects of reliability and criticality on an IED interrogation task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1),* pp. 2340-2344. Chicago, USA:. Sage Publications.

Harris, W. C., Hancock, P. A., Arthur, E. J., & Caird, J. K. (1995). Performance, workload, and fatigue changes associated with automation. *The International Journal of Aviation Psychology*, *5*(2), 169-185.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: Elsevier.

Heereman, J., & Walla, P. (2011). Stress, uncertainty and decision confidence. *Applied psychophysiology and biofeedback*, *36*(4), 273-279.

Hoffman, B., & Spatariu, A. (2008). The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary educational psychology*, *33*(4), 875-893.

Hoffman, R. R., & Klein, G. L. (2017). Challenges and Prospects for the Paradigm of Naturalistic Decision Making. *Journal of Cognitive Engineering and Decision Making*, *11(1)*, 97-104.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human factors*, *40*(2), 254-276.

Hu, Y., Wang, D., Pang, K., Xu, G., & Guo, J. (2015). The effect of emotion and time pressure on risk decision-making. *Journal of Risk Research*, *18*(5), 637-650.

Hubert-Wallander, B., Green, C. S., & Bavelier, D. (2011). Stretching the limits of visual attention: the case of action video games. *Wiley interdisciplinary reviews: cognitive science*, *2*(2), 222-230.

Hunt, (2014). Dstl. Subject matter expert personal communication. Portsdown West. Fareham.

Iannello, P., Mottini, A., Tirelli, S., Riva, S., & Antonietti, A. (2017). Ambiguity and uncertainty tolerance, need for cognition, and their association with stress. A study among Italian practicing physicians. *Medical education online*, *22*(1), 1270009.

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of loss prevention in the process industries*, *32*, 454-465.

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9(1)*, 25-49.

Jackson, S. A., Kleitman, S., & Aidman, E. (2014). Low cognitive load and reduced arousal impede practice effects on executive functioning, metacognitive confidence and decision making. *PloS one*, *9*(12), e115689.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3*(2), 114-158.

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of personality and social psychology*, *45*(1), 20.

Johnston, J. H., Poirier, J., and Smith-Jentsch, K. A. 1998. Decision making under stress: Creating a research methodology. In Cannon-Bowers J.

A. & Salas E. (Eds.). Making Decisions Under Stress: Implications for Individuals and Teams. (39-59). American Psychological Association: Washington, DC.

Jøsok, Ø., Knox, B. J., Helkala, K., Lugo, R. G., Sütterlin, S., & Ward, P. (2016). Exploring the hybrid space. *International Conference on Augmented Cognition, July 2016,* (pp. 178-188). Chicago, USA: Springer International Publishing.

Juanchich, M., Dewberry, C., Sirota, M., & Narendran, S. (2016). Cognitive reflection predicts real-life decision outcomes, but not over and above personality and decision-making styles. *Journal of Behavioral Decision Making*, *29*(1), 52-59.

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision-making in complex naval command-and-control environments. *Human Factors*, 38(2), 220-231.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kebbell, M. R., Muller, D. A., & Martin, K. (2010). Understanding and managing bias. *Dealing with uncertainties in policing serious crime*, *16*(1), 87.

Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology*, *87(4)*, 653-662.

Keinan, G., Friedland, N., & Ben-Porath, Y. (1987). Decision making under stress: Scanning of alternatives under physical threat. *Acta Psychologica*, *64*(3), 219-228.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1594), 1322-1337.

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217-273.

Kerstholt, J.H, & Pieters, A.R. (1994). *A Comparison of Decision Making Behavior in a Static and a Dynamic Task Environment as a Function of Time Pressure (A Comparison of Decision Behavior in a Static* *and a Dynamic Task Environment as a Function of Time Pressure)* (No. TNO-TM-1994-B-10). Human factors research inst tno soesterberg (Netherlands).

Kim, J. H. (2018). The effect of metacognitive monitoring feedback on performance in a computer-based training simulation. *Applied ergonomics*, *67*, 193-202.

Kim, J. H., Macht, G. A., & Li, S. (2012, September). Comparison of Individual and Team-Based Dynamic Decision-Making Task (Anti-Air Warfare Coordinator): Consideration of Subjective Mental Workload Metacognition. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*(Vol. 56, No. 1, pp. 2517-2521). Sage CA: Los Angeles, CA: SAGE Publications.

Klein, G. (1996). The effect of acute stressors on decision making. *Stress and human performance*, 49-88.

Klein, G. (2015). Reflections on applications of naturalistic decision making. *Journal of Occupational and Organizational Psychology*, *88*(2), 382-386.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In *Proceedings of the Human Factors Society Annual Meeting*(Vol. 30, No. 6, pp. 576-580). Sage CA: Los Angeles, CA: Sage Publications.

Klein, G. A., Orasanu, J. E., Calderwood, R. E., & Zsambok, C. E. (1993). Decision making in action: Models and methods. In *This book is an outcome of a workshop held in Dayton, OH, Sep 25–27, 1989.* Ablex Publishing.

Klein, G., & Militello, L. (2001). Some guidelines for conducting a cognitive task analysis. In *Advances in human performance and cognitive engineering research* (pp. 163-199). Emerald Group Publishing Limited.

Klein, G., Hintze, N., & Saab, D. (2013). Thinking inside the box: The ShadowBox method for cognitive skill development. In *Proceedings of the 11th International Conference on Naturalistic Decision Making* (pp. 121-124).

Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. Applied Cognitive Psychology, 15, 321–341.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and individual differences*, *17*(2), 161-173.

Kliegel, M., Martin, M., McDaniel, M., & Einstein, G. (2004). Importance effects on performance in event-based prospective memory tasks. *Memory*, *12(5)*, 553-561.

Kozlowski, S. W., & DeShon, R. P. (2004). A psychological fidelity approach to simulation-based training: Theory, research and principles. *Scaled worlds: Development, validation, and applications*, 75-99.

Lanzetta, J. T. (1963). Information acquisition in decision making. *Motivation and social interaction-cognitive determinants*, 239-265.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

Lee, J. D. (1999, September). Measuring driver adaptation to in-vehicle information systems: Disassociation of subjective and objective situation awareness measures. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, No. 18, pp. 992-996). Sage CA: Los Angeles, CA: SAGE Publications.

Lichacz, F. M. (2008). Augmenting understanding of the relationship between situation awareness and confidence using calibration analysis. *Ergonomics*, *51*(10), 1489-1502.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In H. Jungermann & G. de Zeeuw (Eds.) *Decision Making and Change in Human Affairs* (pp. 275-324). Dordrecht, Netherlands: Reidel.

Liebhaber, M. J., & Smith, C. A. (2000). *Naval air defense threat assessment: Cognitive factors and model*. PACIFIC SCIENCE AND ENGINEERING GROUP INC SAN DIEGO CA.

Lin, J., Wohleber, R., Matthews, G., Chiu, P., Calhoun, G., Ruff, H., & Funke, G. (2015, September). Video game experience and gender as predictors of performance and stress during supervisory control of multiple unmanned aerial vehicles. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 746-750). Sage CA: Los Angeles, CA: SAGE Publications.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69(2)*, 149-163.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14(5)*, 331-352.

Liu, D., Macchiarella, N. D., Vincenzi, D. A., Vincenzi, D. A., Wise, J. A., Mouloua, M., & Hancock, P. A. (2009). Human Factors in Simulation and Training. Orlando, FL:Pamtership 1.

Loft, S., Sadler, A., Braithwaite, J., & Huf, S. (2015). The chronic detrimental impact of interruptions in a simulated submarine track management task. *Human factors*, *57*(8), 1417-1426.

Lovallo, D., & Kahneman, D. (2003). Delusions of success. *Harvard Business Review*, *81(7)*, 56-63.

Maniscalo, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*, 422-430.

Markman, A. B. (2018). Combining the strengths of naturalistic and laboratory decision-making research to create integrative theories of choice. *Journal of applied research in memory and cognition*, *7*(1), 1-10.

Marsh, J. E., Demaine, J., Bell, R., Skelton, F. C., Frowd, C. D., Röer, J. P., & Buchner, A. (2015). The impact of irrelevant auditory facial descriptions on memory for target faces: Implications for eyewitness memory. *Journal of Forensic Practice*, *17*(4), 271-280.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits*. Cambridge University Press.

Maule, A. J., & Edland, A. C. (1997). The effects of time pressure on human judgment and decision making. *Decision making: Cognitive models and explanations*, 189-204.

Maule, A. J., & Svenson, O. (1993). Theoretical and empirical approaches to behavioral decision making and their relation to time constraints. In *Time pressure and stress in human judgment and decision making* (pp. 3-25). Springer, Boston, MA.

Maule, A. J., Hockey, G. R. J., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica*, *104(3)*, 283-301.

Mayseless, O., & Kruglanski, A. W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes*, *39*(2), 162-183.

McAndrew, C., & Gore, J. (2013). Understanding preferences in experience-based choice: A study of cognition in the “wild”. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 179-197.

McGhee, W., Shields, M. D., & Birnberg, J. G. (1978). The effects of personality on a subject's information processing. *Accounting Review*, 681-697.

McGuinness, B. (2004). *Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings*. *Command and Control Research and Technology Symposium: The Power of Information Age Concepts and Technologies,* San Diego, California, 15th – 17th June 2004.

McKinley, R. A., McIntire, L. K., & Funke, M. A. (2011). Operator selection for unmanned aerial systems: comparing video game players and pilots. *Aviation, space, and environmental medicine*, *82*(6), 635-642.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618-1641.

Ministry of Defence (1986). Report of board of inquiry into loss of an army air corps gazelle over the Falkland Islands of 6 June 1982; retrieved January 2015 from, http://webarchive.nationalarchives.gov.uk/20121026065214/http://www.mod.uk/NR/rdonlyres/EF248AAE-5B25-4CB4-BE90-EE096980354B/0/boi_loss_gazellexx377.pdf.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of experimental psychology: General*, *130*(4), 621.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129-140.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher-order dimensions of the big five personality traits and the big six vocational interest types. *Personnel Psychology*, *58*(2), 447-478.

NATO (2008). *Department of defence interface standard*. Received January 2015from http://www.dtic.mil/doctrine//doctrine/other/ms_2525c.pdf..

NATO (2014). AAP-06. *NATO Glossary of terms and definition (English and French*). Received January 2015 from http://wcnjk.wp.mil.pl/plik/file/N_20130808_AAP6EN.pdf.

Nicholson, B., & O'Hare, D. (2014). The effects of individual differences, prior experience and cognitive load on the transfer of dynamic decision-making performance. *Ergonomics*, *57*(9), 1353-1365.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Educational*, 7-28.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231.

O'Brien, K. S., & O'Hare, D. (2007). Situational awareness ability and cognitive skills training in a complex real-world task. *Ergonomics*, *50*(7), 1064-1091.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.), *Decision Making in Action: Models and Methods* (pp. 3-20). Norwood, NJ: Ablex.

Orasanu, J., Fischer, U., McDonnell, L. K., Davison, J., Haars, K. E., Villeda, E., & VanAken, C. (1998, October). How do flight crews detect and prevent errors? Findings from a flight simulation study. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 42, No. 3, pp. 191-195). Sage CA: Los Angeles, CA: SAGE Publications.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of general psychology*, *129*(3), 257-299.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140-160.

Parker, A. M., & Stone, E. R. (2014). Identifying the effects of unjustified confidence versus overconfidence: Lessons learned from two analytic methods. *Journal of Behavioral Decision Making*, *27(2),* 134-145.

Penley, J. A., & Tomaka, J. (2002). Associations among the Big Five, emotional responses, and coping with acute stress. *Personality and Individual Differences*, *32(7)*, 1215-1228.

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic bulletin & review*, *10*(1), 177-183.

Piedmont, R. L., & Weinstein, H. P. (1993). A psychometric evaluation of the new NEO–PIR facet scales for agreeableness and conscientiousness. *Journal of Personality Assessment*, *60*(2), 302-318.

Poos, J. M., van den Bosch, K., & Janssen, C. P. (2017). Battling bias: Effects of training and training context. *Computers & Education*, *111*, 101-113.

Presagia. (2013). *Virtual Avionics Prototyping Software (VAPS) XT Course Notes 4.0.* [PDF file]. Retrieved January 2015.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). Cognitive systems engineering.

Reimer, T., & Rieskamp, J. (2007). Fast and frugal heuristics. *Encyclopedia of social psychology*, 346-348.

Reinerman-Jones, L., Matthews, G., Burke, S., & Scribner, D. (2016, September). A Situation Judgment Test for military multicultural decision-making: Initial psychometric studies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1482-1486). Sage CA: Los Angeles, CA: SAGE Publications.

Reinten, J., Kort, H. S. M., Hornikx, M. C. J., & Kohlrausch, A. G. (2017, June). Context specific analysis of the sound environment at the workplace and its relation with a task. In *12th ICBEN Congress on Noise as a Public Health Problem, 18-22 June, Zurich, Switzerland*.

Roberts, A. P., & Cole, J. C. (2018). Naturalistic Decision Making: Taking a (Cognitive) Step Back to Take Two Steps Forward in Understanding Experience-Based Decisions. *Journal of applied research in memory and cognition*, *7*(1), 70-81.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43(2)*, 137-145.

Rodes, W., & Gugerty, L. (2012). Effects of electronic map displays and individual differences in ability on navigation performance. *Human factors*, *54*(4), 589-599.

Roets, A., & Van Hiel, A. (2007). Separating ability from need: Clarifying the dimensional structure of the need for closure scale. *Personality and Social Psychology Bulletin*, *33(2)*, 266-280.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. IEEE Transactions on Systems, Man & Cybernetics, 22, 1296–1308.

Rousseau, R., Tremblay, S., Banbury, S., Breton, R., & Guitouni, A. (2010). The role of metacognition in the relationship between objective and subjective measures of situation awareness. *Theoretical Issues in Ergonomics Science*, *11(1-2)*, 119-130.

Roux, J. N., & Van Vuuren, J. H. (2007). Threat evaluation and weapon assignment decision support: A review of the state of the art. *ORiON*, *23*(2), 151-187.

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review*, *18*, 46-57.

Salas, E. E., & Fiore, S. M. (2004). *Team cognition: Understanding the factors that drive process and performance*. American Psychological Association.

Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, *11*(2-3), 271-288.

Salmon, P., Jenkins, D., Stanton, N., & Walker, G. (2010). Hierarchical task analysis vs. cognitive work analysis: comparison of theory, methodology and contribution to system design. *Theoretical Issues in Ergonomics Science, 11*(6), 504-531.

Salmon, P., M., Stanton, N., A., Walker, G., H., & Green, D. (2006). Situation Awareness Measurement: A Review of Applicability for C4i environments. Applied Ergonomics, 37(1), 225-238.

Saus, E. R., Johnsen, B. H., Eid, J., & Thayer, J. F. (2012). Who benefits from simulator training: Personality and heart rate variability in relation to situation awareness during navigation training. *Computers in Human Behavior*, *28*(4), 1262-1268.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K.(2004). Overconfidence and the Big Five. Journal of Research in Personality, 38(5), 473−480.

Schaeffer, M. H. (1989). Environmental stress and individual decision-making: Implications for the patient. *Patient Education and Counseling*, *13*(3), 221-235.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology, 6*, 147–151.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and learning*, *4*(1), 33-45.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology*, *19*(4), 460-475.

Selcon, S. J., Taylor, R. M., & Koritsas, E. (1991). Workload or situational awareness? TLX vs. SART for aerospace systems design evaluation. *Proceedings of the Human Factors Society Annual Meeting, 35(2),* September 1991, pp. 62-66. Sage CA: Los Angeles, CA: SAGE Publications.

Sieber, J. E., & Lanzetta, J. T. (1964). Conflict and conceptual structure as determinants of decision-making behaviour. *Journal of Personality*, *32*(4), 622-641.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69(1)*, 99-118.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational behavior and human decision processes*, *43*(1), 1-28.

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and cognition, 20*(4), 1787-1792.

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods, 19*(3), 506-532.

Spector, P. E., & O'Connell, B. J. (1994). The contribution of personality traits, negative affectivity, locus of control and Type A to the subsequent reports of job stressors and job strains. *Journal of Occupational and Organizational psychology*, *67*(1), 1-12.

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92.

St John, M., Callan, J., Proctor, S., & Holste, S. T. (2000). *Tactical decision-making under uncertainty:* Experiments I and II (Technical. Arlington, CA: Office of Naval Research.

Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, *100(4),* 961.

Stankov, L., Kleitman, S., & Jackson, S. A. (2015). Measures of the trait of confidence. In *Measures of personality and social psychological constructs* (pp. 158-189).

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety?. *Learning and Individual Differences*, *22*(6), 747-758.

Stanners, M., & French, H. T. (2005). *An empirical study of the relationship between situation awareness and decision making* (No. DSTO-TR-1687). Defence Science and Technology Organization Edinburgh (Australia) Land Operations Div.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. (2005). Human factors methods: A practical guide for engineering and design. Ashgate: Aldershot.

Stichman, E. P. (1967). *Relationship of Expressed Confidence to Accuracy of Transcription by Operational Communications Personnel* (No. BESRL-TRN-192). Army Behavoir and Systems Research Lab Arlington Va.

Sulistyawati, K., & Chui, Y. P. (2009). Confidence Bias in Situation Awareness. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 317-325). Springer, Berlin, Heidelberg.

Sulistyawati, K., Wickens, C. D., & Chui, Y. P. (2011). Prediction in situation awareness: Confidence bias and underlying cognitive abilities. *The International Journal of Aviation Psychology*, *21*(2), 153-174.

Szalma, J. L. (2009). Individual differences in human-technology interaction: Incorporating variation in human characteristics into human factors and ergonomics research and design. *Theoretical Issues in Ergonomics Science*, 10, 381–397.

Tanner Jr, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological review*, *61*(6), 401.

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11*(2), 113-120.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *In: AGARD Conference Proceedings No 478, Situational Awareness in Aerospace Operations. Aerospace Medical Panel Symposium*, (AGARD-CP-478) pp3/1 –3/17, Neuilly Sur Seine, France: NATO-AGARD.

Teng, C. I. (2009). Online game player personality and real-life need fulfillment. *International Journal of Cyber Society and Education*, *2*(2), 39-50.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current directions in psychological science*, *16*(3), 167-171.

Turkle, S. (1984). *The second self: Computers and the human spirit*. London, Granada.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Vachon, F., Tremblay, S., Nicholls, A. P., & Jones, D. M. (2011, September). Exploiting the auditory modality in decision support: Beneficial “warning” effects and unavoidable costs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 1402-1406). Sage CA: Los Angeles, CA: Sage Publications.

van den Heuvel, C., Alison, L., & Crego, J. (2012). How Uncertainty and Accountability can Derail Strategic 'Save Life'Decisions in Counter-Terrorism Simulations: A Descriptive Model of Choice Deferral and Omission Bias. *Journal of Behavioral Decision Making*, *25*(2), 165-187.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of human factors and ergonomics*, *4*, 243-273.

Vidulich, M. A., Crabtree, M. S., & McCoy, A. L. (1993). Developing subjective and objective metrics of pilot situation awareness. In *Proceedings of the 7th International Symposium on Aviation Psychology* (pp. 896-900).

Vidulich, M. A., McCoy, A. L., & Crabtree, M. S. (1995). Attentional control and situational awareness in a complex air combat simulation. *Situation awareness: Limitations and enhancement in the aviation environment*, 18.

Walker, M., Takayama, L., & Landay, J. A. (2002, September). High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 46, No. 5, pp. 661-665). Sage CA: Los Angeles, CA: SAGE Publications.

Wallenius, C., Bäccman, C., & Larsson, G. (2014). Examining the impact of personality and situational factors on decision making among military staffs. *Journal of Military Studies*, *5*(2), 26-37.

Ward, P., Gore, J., Hutton, R., Conway, G. E., & Hoffman, R. R. (2018). Adaptive skill as the conditio sine qua non of expertise. *Journal of applied research in memory and cognition*.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of personality and social psychology*, *67*(6), 1049.

Wheatcroft, J. M., & Woods, S. (2010). Effectiveness of Witness preparation and cross-examination non-directive and directive leading question styles on witness accuracy and confidence. *The International Journal of Evidence & Proof*, *14(3)*, 187-207.

Wheatcroft, J. M., Jump, M., Breckell, A. L., & Adams-White, J. (2017). Unmanned aerial systems (UAS) operators' accuracy and confidence of decisions: Professional pilots or video game players? *Cogent Psychology, 4(1)*.

Wheatcroft, J.M., Wagstaff, G.F & Kebbell, M.R. (2004). The influence of courtroom questioning style on actual and perceived eyewitness confidence and accuracy. *Legal & Criminological Psychology, 9*, 83-101.

Wheatcroft, J.M., Wagstaff, G.F., & Manarin, B. (2015). The influence of delay and item difficulty on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research, 1,* 1-9.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current directions in psychological science*, *11*(4), 128-133.

Wiggins, M., & O'hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, *1*(4), 305.

Williams, P. G., Rau, H. K., Cribbet, M. R., & Gunn, H. E. (2009). Openness to experience and stress regulation. *Journal of Research in Personality*, *43(5)*, 777-784.

Wittek, C. T., Finserås, T. R., Pallesen, S., Mentzoni, R. A., Hanss, D., Griffiths, M. D., & Molde, H. (2016). Prevalence and predictors of video game addiction: A study based on a national representative sample of gamers. *International journal of mental health and addiction*, *14*(5), 672-686.

Yang, H., Thompson, C., & Bland, M. (2012). The effect of clinical experience, judgment task difficulty and time pressure on nurses' confidence calibration in a high fidelity clinical simulation. *BMC medical informatics and decision making*, *12*(1), 113.

Yates, J. F., & Tschirhart, M. D. (2006). Decision-making expertise. *The Cambridge handbook of expertise and expert performance*, 421-438.

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.

Zsambok, C. E., & Klein, G. A. (Eds.) (1997). *Naturalistic decision making*. Mahwah, N.J: Lawrence Erlbaum Associates.

Appendices

Appendix 1a: The University of Liverpool ethic approval confirmation

IPHS-1516-21-Investigating Decision Accuracy and Decision Confidence in Human-Machine Interaction within a Ships Operations Room. (Investigating Decision Making in Navy Ships Operations Rooms)

IPHS Ethics

Sent: 21 October 2015 16:07

To: Wheatcroft, Jacqueline

Dear Jacqueline

I am pleased to inform you that IPHS Research Ethics Committee has approved your application for ethical approval. Details and conditions of the approval can be found below.

Ref: IPHS-1516-21

PI / Supervisor: Jacqueline Wheatcroft

Title: Investigating Decision Accuracy and Decision Confidence in Human-Machine Interaction within a Ships Operations Room. (Investigating Decision Making in Navy Ships Operations Rooms)

First Reviewer: Charlotte Hardman

Second Reviewer: Georg Meyer

Date of Approval: 21.10.15

The application was APPROVED subject to the following conditions:

Conditions

1 All serious adverse events must be reported to the Sub-Committee within 24 hours of their occurrence, via the Research Governance Officer (ethics@liv.ac.uk).

2 This approval applies for the duration of the research. If it is proposed to extend the duration of the study as specified in the application form, IPHS REC should be notified as follows. If it is proposed to make an amendment to the research, you should notify IPHS

REC by following the Notice of Amendment procedure outlined at http://www.liv.ac.uk/researchethics/amendment%20procedure%209-08.doc.

3 If the named PI / Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore please contact the Institute's Research

Ethics Office at iphsrec@liverpool.ac.uk in order to notify them of a change in PI / Supervisor.

Best Wishes

Liz Brignal

Secretary, IPHS Research Ethics Committee

Email: iphsrec@liv.ac.uk

Appendix 1b: The Ministry of Defence Research Ethical approval confirmation

Ministry of Defence
Research Ethics Committee

MODREC Secretariat
Building 5, GO2,
Defence Science and Technology Laboratory,
Porton Down, Salisbury, SP4 0JQ

Telephone: 01980 956351
e-mail: MODREC@dstl.gov.uk

Dr Jacqueline Wheatcroft
University of Liverpool,
Institute of Psychology,
Health & Society,
Department of Psychological Sciences,
Eleanor Rathbone Building,
Bedford Street South,
Liverpool
L69 7ZA

Telephone: 0151 7950513

E-mail: jacmw@liverpool.ac.uk

Our Reference: 759/MODREC/16

Date: 26th January 2017

Dear Jacqueline,

Investigating Decision Accuracy and Decision Confidence in Human-Machine Interaction within a Ships Operations Room. (Investigating Decision Making in Navy Ships Operations Rooms)

Thank you for submitting your revised Protocol 759/MODREC/16 with tracked changes, and the covering letter with detailed responses to the MODREC letter. I can confirm the revised protocol has been given favourable opinion ex-Committee, and wish you and your colleagues a successful study.

In due course please send the Secretariat a final report containing a summary of the results, so this can be filed in accordance with MODREC arrangements. Please would you also send a brief interim report in one year if the study is still on-going at that point.

This approval is valid for three years and is conditional upon adherence to the protocol – please let me know if any amendment becomes necessary.

Yours sincerely

Ministry of Defence
Research Ethics Committee

MODREC Secretariat
Building 5, GO2,
Defence Science and Technology Laboratory,
Porton Down, Salisbury, SP4 0JQ

Telephone: 01980 956351
e-mail: MODREC@dstl.gov.uk

Dr Jacqueline Wheatcroft
University of Liverpool,
Institute of Psychology,
Health & Society,
Department of Psychological Sciences,
Eleanor Rathbone Building,
Bedford Street South,
Liverpool
L69 7ZA

Telephone: 0151 7950513

E-mail: jacmw@liverpool.ac.uk

Our Reference: 759/MODREC/16

Date: 22 March 2017

Dear Jacqueline,

Investigating Decision Accuracy and Decision Confidence in Human-Machine Interaction within a Ships Operations Room. (Investigating Decision Making in Navy Ships Operations Rooms)

You have requested two amendments to the subject application:

a. The addition of an additional time pressure dimension to the research.
b. The addition of an additional audio element to the research.

You have explained that you wish to extend the programme to examine the effect of time pressure on participant's decision making abilities (i.e., those already included in the approved protocol). Instead of participants being allowed 20 seconds to make a decision (i.e., current protocol), participants will be allowed 10 seconds to make decisions.

You further wish to extend the programme to examine the effect of audio communication on participant's decision making abilities (i.e., those already included in the approved protocol). Half the participants will be instructed that they must attend to the information they hear during the task as there will be a series of questions that will be asked of them at the end of the scenario. The other half will be advised they will hear background noise and it is not necessary for them to attend to it.

You have advised that the student is in her third year and is struggling to obtain participants. You said you have been advised it will be difficult to obtain 30 MOD participants and you have therefore taken the view that further studies need to give the student a chance to be able to collect data. You want to leave in the MOD participants (as prospects) but have no realistic expectation of being able to obtain more than 30.

The sub-committee have given approval to the stated amendments.

Yours sincerely

Dr Simon Kolstoe,
Chair, MODREC

Appendix 2a: SART

Situation Awareness is defined as "timely knowledge of what is happening as you perform tasks during the mission."

	Situation Awareness Rating Techniques (SART)	
Demand	Instability of Situation	Likeliness of situation to change suddenly
	Variability of Situation	Number of variables which require your attention
	Complexity of Situation	Degree of complication(number of closely connected parts) of the situation
Supply	Arousal	Degree to which you are ready for activity
	Spare Mental Capacity	Amount of mental ability available to apply to new tasks
	Concentration	Degree to which your thoughts are brought to bear on the situation
	Division of Attention	Amount of division of your attention in the situation
Understanding	Information Quantity	Amount of knowledge received and understood
	Information Quality	Degree of goodness or value of knowledge communicated
	Familiarity	Degree of acquaintance with the situation

Rate the level of each component of situation awareness that you had when you performed the tasks during the mission that you just completed. Circle the appropriate number for each component of situation awareness (e.g., complexity of situation).

DEMAND

Instability of situation: Low 1---------2---------3---------4---------5---------6---------7 High

Variability of situation: Low 1---------2---------3---------4---------5---------6---------7 High

Complexity of situation: Low 1---------2---------3---------4---------5---------6---------7 High

SUPPLY

Arousal: Low 1---------2---------3---------4---------5---------6---------7 High

Spare mental capacity: Low 1---------2---------3---------4---------5---------6---------7 High

Concentration: Low 1---------2---------3---------4---------5---------6---------7 High

Division of attention: Low 1---------2---------3---------4---------5---------6---------7 High

UNDERSTANDING

Information quantity: Low 1---------2---------3---------4---------5---------6---------7 High

Information quality: Low 1---------2---------3---------4---------5---------6---------7 High

Familiarity: Low 1---------2---------3---------4---------5---------6---------7 High

Appendix 2b: NASA TLX: Pairwise comparison cards & rating sheet

Effort Or Performance	Temporal Demand Or Frustration
Temporal Demand Or Effort	Physical Demand Or Frustration
Performance Or Frustration	Physical Demand Or Temporal Demand
Physical Demand Or Performance	Temporal Demand Or Mental Demand

Frustration Or Effort	Performance Or Mental Demand
Performance Or Temporal Demand	Mental Demand Or Effort
Mental Demand Or Physical Demand	Effort Or Physical Demand
Frustration Or Mental Demand	

Participant Number: ______________________________

RATING SHEET

MENTAL DEMAND

Low **High**

PHYSICAL DEMAND

Low **High**

TEMPORAL DEMAND

Low **High**

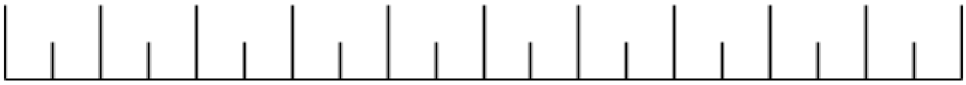
PERFORMANCE

Good **Poor**

EFFORT

Low **High**

FRUSTRATION



Low **High**

Appendix 3: Consent form

Committee on Research Ethics

PARTICIPANT CONSENT FORM

Title of Research Project:	**Investigating Decision Making in Navy Ship's Operations Rooms**
Researcher(s):	Jade Adams-White

1. I confirm that I have read and have understood the information sheet dated [01/09/17] for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my rights being affected. In addition, should I not wish to answer any particular question or questions, I am free to decline. ☐

3. I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my anonymised responses. I understand that neither my name nor affiliation will be linked with the research materials, nor, will I be identified or identifiable in the reports or outputs that result from the research. ☐

4. I understand that, under the Data Protection Act, I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish. ☐

5. I agree to take part in the above study. ☐

6. I agree, that should I withdraw, data already provided can be used by the researchers. ☐

Participant Name	Date	Signature
____________________	________	____________________

Researcher Jade Adams-White	Date	Signature
____________________	________	____________________

Principal Investigators:
Dr Jacqueline Wheatcroft
University of Liverpool
Department of Psychological Sciences

Student Researcher:
Jade Adams-White
University of Liverpool
School of Engineering

Appendix 4a: Participant information sheet – Students and VGPs

Study title

Investigating Decision Making in Navy Ship's Operations Rooms

Invitation to take part

You are being invited to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask the researcher if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends and relatives if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

Thank you for reading this.

What is the purpose of the research?

The research is being conducted by Jade Adams -White, who is a PhD student at the University of Liverpool. The research is being conducted as part of the requirements for completing the Doctorate in Psychology.

The purpose of this study is to increase understanding of factors that influence and contribute to decision accuracy and confidence in air defence decision making. There is potential for this research to be used to help prioritize training and individual needs, and selection, in order to improve the effectiveness of decision making in air defence and determine how decision support tools might best be used.

Who is doing this research?

The University of Liverpool (Dr Jacqueline Wheatcroft, Dr Mike Jump, and Jade Adams-White), in conjunction with Defence Science and Technology Laboratory (Dstl).

Why have I been invited to take part?

You have been chosen as you are a student.

Do I have to take part?

No. Your participation in the research study is entirely voluntary, and you are free to withdraw at any time without explanation and without incurring any disadvantage.

What will I be asked to do?

You will be seated in front of a monitor and be provided with the participant information sheet to read. You will then be asked if you have any questions and once satisfied you have a good understanding of what it is the study requires of you , and you have agreed to participate, you will be asked to sign the consent form in order to proceed. You will first complete participant demographic form which collects data on age, gender and occupation The type/s of game/s played, together with the average time spent playing games a week will also be collected. You will then be asked to complete paper based questionnaires to gauge the relevance of a number of measures (e.g., general personality constructs, thinking and reasoning). Following this, you will be provided with the task booklet to read. The task booklet provides you with information needed to assist you in the decision making task, including air defence terminology and symbols. Once you have read the booklet you will undertake a practice trial.

Once happy you understand the task you will take part in the experimental condition. For this, you will observe an air defence scenario playing out on the monitor. The scenario will pause at certain time intervals and you will be required to make a decision and rate your confidence in that decision in a questionnaire booklet. The booklet presents three (3) separate decision options based on the events of the scenario. One choice is required to be selected by placing a tick by the option you believe to be the 'best option given the current situation'. You will then be required to rate how confident you are in the options chosen on a Likert scale, where 0 = not at all confident to 5 = extremely confident.

You will also complete Situational Awareness and Workload and Visual Analogue Mood questionnaires.

The study will last between 60 – 90 minutes.

What are the benefits of taking part?

Year 1 UoL psychology students can be awarded EPR points for taking part in experiments.

What are the possible disadvantages and risks of taking part?

Given the nature of the study no adverse risks or disadvantages are anticipated.

Can I withdraw from the research and what will happen if I don't want to carry on?

You have the right to withdraw your data at any time, without explanation. Results up to the period of withdrawal may be used, if you are happy for this to be done. Otherwise you may request that they are destroyed and no further use is made of them. As results will be anonymised they may only be withdrawn prior to the anonymisation process being completed.

Are there any expenses and payments which I will get?

There are no payments for taking part in the research.

Will my taking part or not taking part affect my Service career or medical care?

No

Whom do I contact if I have any questions or a complaint?

If you are unhappy, or if there is a problem, please feel free to let us know by contacting the main researcher (Jade Adams-White) or by contacting Jacqueline Wheatcroft at the University of Liverpool (jacmw@liverpool.ac.uk / 01517950513) and we will try to help. If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Governance Officer at ethics@liv.ac.uk. When contacting the Research Governance Officer (i.e., Participant Advocate), please provide details of the name or description of the study (so that it can be identified), the researcher involved, and the details of the complaint you wish to make.

What happens if I suffer any harm?

Given that the research will take place in a normal office environment either at the university or service locations, it is not anticipated that any serious safety events are likely to occur. Any event that is a cause for concern for the research team member will be reported to the Principal Investigator/Safety Officer and all will be made aware of their responsibilities for reporting any safety issues that arise during the project

There are unlikely to be any adverse physical or psychological effects of participation and so the risks are minimal; participants are not expected to experience any lasting effects. Although it is not anticipated that adverse effects will occur, if they do, the study will be paused immediately and the problems will be reported to the Ethics sub-committee within 24 hours of their occurrence through the Research Governance Officer (ethics@liverpool.ac.uk).

Please see the consent form for further details.

Will my records be kept confidential?

All information provided will be kept strictly confidential and will be anonymised, including your name and any affiliations. Data files will be password protected and stored in the University's secure server. Random assignment number codes will be applied by the researcher, in order to ensure that data is anonymised. Data collected will be used for the purposes of the project at University of Liverpool. Only the named researcher and supervisors will have access to the raw data, which will remain stored for 50 years in the University's secure server, and then deleted.

Who is organising and funding the research?

Organiser : University of Liverpool
Funder : Defence Science and Technology Laboratory (Dstl)

Who has reviewed the study?

The study has been reviewed and given a favourable ethical opinion by the Ministry of Defence Research Ethics Committee (MoDREC)

Further information and contact details.

Please contact:

Principal Investigators:	**Student Researcher:**
Dr Jacqueline Wheatcroft	Jade Adams-White
University of Liverpool	University of Liverpool
Department of Psychological Sciences	School of Engineering
01517950513	0151944814
jacmw@liverpool.ac.uk	j.adams-white@liverpool.ac.uk

Mike Jump

University of Liverpool

School of Engineering

Mjump1@liverpool.ac.uk

Compliance with the Declaration of Helsinki.

This study complies, and at all times will comply, with the Declaration of Helsinki[1] as adopted at the 64th WMA General Assembly at Fortaleza, Brazil in October 2013.

[1] World Medical Association Declaration of Helsinki [revised October 2013]. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. 64th WMA General Assembly, Fortaleza (Brazil).

Appendix 4b: Participant information sheet – Military

Study title

Investigating Decision Making in Navy Ship's Operations Rooms

Invitation to take part

You are being invited to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask the researcher if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends and relatives if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

Thank you for reading this.

What is the purpose of the research?

The research is being conducted by Jade Adams -White, who is a PhD student at the University of Liverpool. The research is being conducted as part of the requirements for completing the Doctorate in Psychology.

The purpose of this study is to increase understanding of factors that influence and contribute to decision accuracy and confidence in air defence decision making. There is potential for this research to be used to help prioritize training and individual needs, and selection, in order to improve the effectiveness of decision making in air defence and determine how decision support tools might best be used.

Who is doing this research?

The University of Liverpool (Dr Jacqueline Wheatcroft, Dr Mike Jump, and Jade Adams-White), in conjunction with Defence Science and Technology Laboratory (Dstl).

Why have I been invited to take part?

You have been chosen as you are a student recruit of the Royal Navy or you are serving (or ex-service) Royal Navy personnel who has experience of Operation Rooms.

Do I have to take part?

No. Your participation in the research study is entirely voluntary, and you are free to withdraw at any time without explanation and without incurring any disadvantage.

What will I be asked to do?

You will be seated in front of a monitor and be provided with the participant information sheet to read. You will then be asked if you have any questions and once satisfied you have a good understanding of what it is the study requires of you , and you have agreed to participate, you will be asked to sign the consent form in order to proceed. You will first complete participant demographic form which collects data on age, gender and information regarding rank and years in the role. The type/s of game/s played, together with the average time spent playing games a week will also be collected. You will then be asked to complete paper based questionnaires to gauge the relevance of a number of measures (e.g., general personality constructs, thinking and reasoning). Following this, you will be provided with the task booklet to read. The task booklet provides you with information needed to assist you in the decision making task, including air defence terminology and symbols. Once you have read the booklet you will undertake a practice trial.

Once happy you understand the task you will take part in the experimental condition. For this, you will observe an air defence scenario playing out on the monitor. The scenario will pause at certain time intervals and you will be required to make a decision and rate your confidence in that decision in the questionnaire booklet. The booklet presents three (3) separate decision options based on the events of the scenario. One choice is required to be selected by placing a tick by the option you believe to be the 'best option given the current situation'. You will then be required to rate how confident you are in the options chosen on a Likert scale, where 0 = not at all confident to 5 = extremely confident.

You will also complete a Situational Awareness and Workload and Visual Analogue Mood questionnaires.

The study will last between 60 – 90 minutes.

Briefings will be held at the shore based establishments in or near to where the participants work, so likely to be Dstl, HMS Collingwood, HMS Nelson and Dryad Maritime.

What are the benefits of taking part?

There are no direct benefits to military participants.

What are the possible disadvantages and risks of taking part?

Given the nature of the study no adverse risks or disadvantages are anticipated.

Can I withdraw from the research and what will happen if I don't want to carry on?

You have the right to withdraw your data at any time, without explanation. Results up to the period of withdrawal may be used, if you are happy for this to be done. Otherwise you may request that they are destroyed and no further use is made of them. As results will be anonymised they may only be withdrawn prior to the anonymisation process being completed.

Are there any expenses and payments which I will get?

There are no payments for taking part in the research.

Will my taking part or not taking part affect my Service career or medical care?

No

Whom do I contact if I have any questions or a complaint?

If you are unhappy, or if there is a problem, please feel free to let us know by contacting the main researcher (Jade Adams-White) or by contacting Jacqueline Wheatcroft at the University of Liverpool (jacmw@liverpool.ac.uk / 01517950513) and we will try to help. If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Governance Officer at ethics@liv.ac.uk. When contacting the Research Governance Officer (i.e., Participant Advocate), please provide details of the name or description of the study (so that it can be identified), the researcher involved, and the details of the complaint you wish to make.

What happens if I suffer any harm?

Given that the research will take place in a normal office environment either at the university or service locations, it is not anticipated that any serious safety events are likely to occur. Any event that is a cause for concern for the research team member will be reported to the Principal Investigator/Safety Officer and all will be made aware of their responsibilities for reporting any safety issues that arise during the project.

There are unlikely to be any adverse physical or psychological effects of participation and so the risks are minimal; participants are not expected to experience any lasting effects. Although it is not anticipated that adverse effects will occur, if they do, the study will be paused immediately and the problems will be reported to the Ethics sub-committee within 24 hours of their occurrence through the Research Governance Officer (ethics@liverpool.ac.uk).

Please see the consent form for further details.

Will my records be kept confidential?

All information provided will be kept strictly confidential and will be anonymised, including your name and any affiliations. Data files will be password protected and stored in the University's secure server. Random assignment number codes will be applied by the researcher, in order to ensure that data is anonymised. Data collected will be used for the purposes of the project at University of Liverpool. Only the named researcher and supervisors will have access to the raw data, which will remain stored for 50 years in the University's secure server, and then deleted.

Who is organising and funding the research?

Organiser : University of Liverpool
Funder : Defence Science and Technology Laboratory (Dstl)

Who has reviewed the study?

The study has been reviewed and given a favourable ethical opinion by the Ministry of Defence Research Ethics Committee (MoDREC)

Further information and contact details.

Please contact:

Principal Investigators:

Dr Jacqueline Wheatcroft

University of Liverpool

Department of Psychological Sciences

01517950513

jacmw@liverpool.ac.uk

Mike Jump

University of Liverpool

School of Engineering

Mjump1@liverpool.ac.uk

Student Researcher:

Jade Adams-White

University of Liverpool

School of Engineering

0151944814

j.adams-white@liverpool.ac.uk

Appendix 5a: Demographics Form: Students and Video Game Players

Participant number: ☐ Age: ☐ Gender: Male/Female

Occupation:

Do you play computer games? Yes/no

A) If yes, on average how many hours a week do you play?

0-5 5-10 10-15 15-20 20-25

B) If yes, what type of computer game do you play most often?

Action (Platform games, Shooter, Fighting)	☐
Role-playing (Fantasy)	☐
Strategy (War games, Real-time Tactics)	☐
Adventure (Stealth, Survival, Horror)	☐
Simulation (Construction & Management, Life, Vehicle)	☐
Sports (Racing, Sports)	☐
Other (please specify)	☐

Appendix 5b: Demographics Form: Military Personnel

Participant number: [ ] Age: [ ] Gender: Male/Female

1. Current Job role: Rank: Length of time in this role: Total Years at sea :

Please list your previous roles, rank and years in that role:

2. Job role: Rank: years in role:

3. Job role: Rank: years in role:

4. Job role: Rank: years in role:

- Do you play computer games? Yes/no

A) If yes, on average how many hours a week do you play?

5-10 10-15 15-20 20-25

B) If yes, on average, do you play 7 or more hours per week and have done over the past 2 years yes/no

C) If yes, what type of computer game do you play most often? **PICK ONE**

Action (Platform games, Shooter, Fighting)	☐
Role-playing (Fantasy)	☐
Strategy (War games, Real-time Tactics)	☐
Adventure (Stealth, Survival, Horror)	☐
Simulation (Construction & Management, Life, Vehicle)	☐
Sports (Racing, Sports)	☐
Other (please specify)	☐

Appendix 6: Debrief

UNIVERSITY OF LIVERPOOL

Committee on Research Ethics

DEBRIEF

Investigating Decision Making in Navy Ships Operations Rooms

Thank you for taking part in the researcher's PhD project. The purpose of this study was to improve our understanding of decision making in Navy Ships Operations Rooms (Ops Room).

The Ops Room is the focal point of the ship. Ops Room personnel deal with vast amounts of information, much of which is incomplete and ambiguous. Thus, decisions made in this environment are challenging and often coupled with time pressures, uncertainty and stress. The research is interested in understanding decision accuracy and confidence across different conditions.

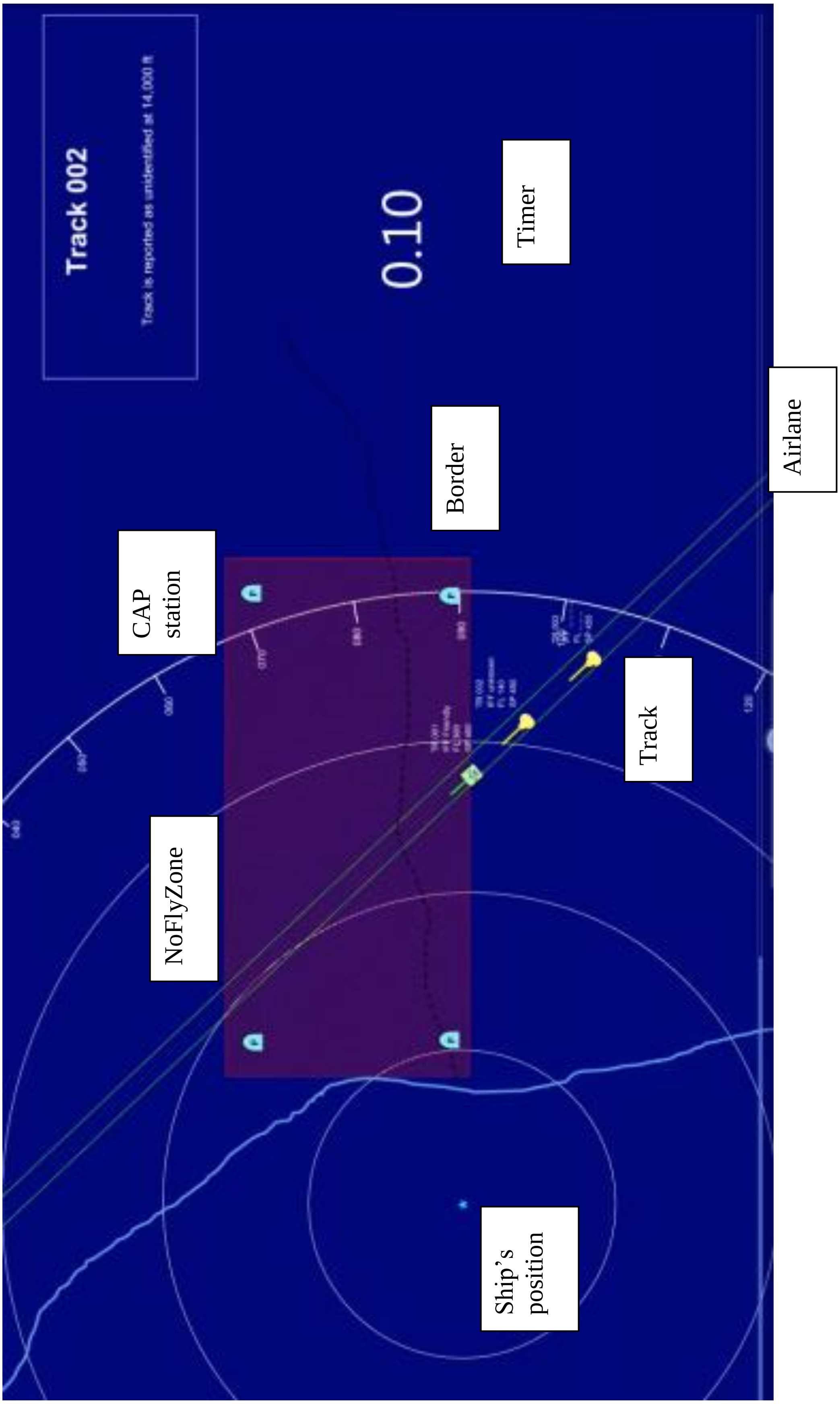
In this study, participants were asked to take part in a scenario and make decisions in response to a range of specific conditions according to peacetime environment. This research will assess the impact of those environments and conditions on ability to make optimum decisions. Participants were allocated to either a low, moderate or high task stress condition and the decisions criticality levels throughout the scenario were varied.

All data collected in this study will be analysed in an aggregated form (your responses will not be singled out); only averaged results will be reported in any future publications. You will remain anonymous.

Thank you again for your participation and helping with this research.

If you would like more information, or have any further questions about any aspect of the study, please contact: Jade Adams-White (j.adams-white@liverpool.ac.uk).

Appendix 7: Scenario Screen Shot



Appendix 8a: Decision Log

Track ID	Event	Decision criticality	Decision option 1	Decision option2	Decision option 3
001	The ship picks up information of a new data link outside the radar. The information is consistent with a civil aircraft following airlane	Low	Examine the link	Confirm that the track is following airlane	Conduct an air investigate procedure by examining all the information from all sensors
001	A sensor is sent from the ship on the bearing of the incoming track.	Medium	Associate the sensor with the link track	Correlate the sensor with the link track	Analyse the sensor parameters for classification
	The sensor bearing remains consistent with the link track and parameters consistent with civair	Medium	Ignore new information	Correlate sensor information with background traffic	Correlate the sensor information with link track
001	A new track appears inside the ships' radar in the vicinity of the incoming link track already located by the sensors. The radar track is transmitting friendly identification friendly or foe (IFF)	Low	Ascertain contact altitude	Do Nothing	Accept IFF identity
002	A new data link appears on the screen -	Low	Examine the information that is presented to you	Confirm that the track is following airlane	Conduct an air investigate procedure by examining all the information from all sensors
003	A new data link appears on the screen -	Low	Examine sensor contribution	Call investigate	Monitor track

002	Track appears on long range radar – no altitude information	Low	Examine sensor contributions	Conduct an air investigate procedure by examining all the information from all sensors	Monitor track
002	IFF information does not correlate with existing track	Medium	Conduct an air investigate procedure by examining all the information from all sensors	Engage	Send the combat fighter patrol aircraft to investigate the track
002	Unidentified track reported 15, 000' altitude	Medium	Engage	Send the combat fighter patrol aircraft to investigate the track	Attempt to establish communication and issue warnings. Warning 1
003	Track appears on long range radar – no altitude information	Low	Examine sensor contribution	Call investigate	Monitor track
001	The track continues in the airlane and starts to move over the NFZ	Low	Do nothing	Advise combat aircraft patrol of assumed civair complying with airlane	Continue to monitor
003	IFF information does not correlate with existing track.	Medium	Conduct an air investigate procedure by examining all the information from all sensors	Engage	Call CAP to investigate

003	Unidentified track reported 15, 000' alt	Medium	Engage	Fighter CAP to investigate	Attempt to establish communication s and issue warnings Warning 1
001	As track 001 crosses over the national border it drops in altitude to 35, 000 ft	Low	Do nothing	Continue to monitor and attempt to contact air traffic control.	Advise CAP in change of flight level
002	Track 002 Enters no fly zone	High	Order CAP to intercept and visual ID.	Continue to read warnings	Order CAP to engage
002	Fighter ID aircrafts as a Country S FGA : visual ID	High	Order the fighter to verify weapon status	Order CAP engagement	Order fighter to attempt communication and read warnings Warning 4
002	Fighter reports that the aircraft has weapons	High	Verify ROE for engagement[2]	Order CAP to establish an air to air position that will allow engagement	Order CAP to engage
003	Track 003 Enters no fly zone	High	Order CAP to intercept and visual ID.	Continue to read warnings	Order CAP to engage
003	Fighter ID aircrafts as a Country S FGA : visual ID	High	Order the fighter to verify weapon status	Order CAP engagement	Order fighter to attempt communication and read warnings Warning 4
003	Fighter reports that the aircraft has weapons	High	Verify ROE for engagement[3]	Order CAP to establish an air to air position that will allow engagement	Order CAP to engage
002	FGA gradually alters course towards the ship .No change in	High	Order CAP to initiate an engagement on track 002	Order CAP to engage	Order CAP to continue to read warnings.

[2] ROE for engagement are: Hostile Act or VIZ ID armed aircraft failing to respond to warnings and operating in the NFZ and/or Self Defence. Collateral damage Low

[3] ROE for engagement are: Hostile Act or VIZ ID armed aircraft failing to respond to warnings and operating in the NFZ and/or Self Defence. Collateral damage Low

	altitude. No communications		up to but not including the point of firing		Warning 4
003	FGA gradually alters course to port. No change in altitude. No communications	Medium	Order CAP to initiate an engagement on a specified track up to but not including the point of firing	Order CAP to engage	Order CAP to continue to read warnings Warning 4
002	FGA leaves no fly zone heading south	Medium	Order CAP to continue to initiate an engagement on track 002 up to but not including the point of firing	Order CAP to shadow from inside no fly zone	Order CAP to discontinue approach but maintain contact unless otherwise indicated.
003	FGA steadies on a heading toward Maritime Task Group. No change in altitude 130 nm	Medium	Order CAP to engage	Order Air Threat Warning Red Attack on the ship is imminent	Order Cap to initiate an engagement on a specified track up to but not including the point of firing and continue to read warnings Warning 4
003	FGA descends to low level and continues to close the task group ('feet dry' - fly's over land)	High	Order CAP to engage	Order Air Threat Warning Red Attack on the ship is imminent	Order Cap to mark and continue to read warnings Warning 5
003	FGA descends to low level and continues to close	High	Engage with force weapons	Cover with force weapons and	Read warnings and monitor for Hostile

	the task group ('feet wet' – fly's over water) 50 nm			Order CAP to discontinue approach but maintain contact unless otherwise indicated.	Intent (ESM electronic support measure) . Set Ship missiles on target Self-defence warning
003	FGA continues to close the Maritime Task Group and demonstrates Hostile Intent 10 nm	High	Continue to read warnings Warning 5	Order CAP to engage.	Verify that CAP is clear and Engage with Force weapons
002	FGA turns towards NFZ again	Medium	Order CAP to read warnings. Warning 2	engage	Order CAP to establish an air-to-air position that will allow engagement of track 002
004	A new data link appears on the screen -	Low	Examine link track data	Validate track is following airlane	Conduct Air investigate procedure.
004	Track appears on long range radar – no altitude information	Low	Examine sensor contribution	Call investigate	Monitor track

Appendix 8b: Task Booklet

DECISION MAKING in NAVY SHIPS OPERATIONS ROOMS

This short booklet has been designed to assist you with your decision making in this task.

Please take your time to read the information presented to you and familiarise yourself with its contents.

You can refer back to this booklet at any time during the task.

SCREEN LAYOUT

A radar will be displayed to you on the screen in front of you. The radar has a radius of 200 Nautical Miles in distance. For the purpose of this task the radar screen has been 'zoomed in'.

On the screen you will see:

- A coastline
- The border between Country S and Country H.
- A designated no fly zone (NFZ)
- An air lane
- Your ship's position
- Textbox which may display information
- Data links/Air tracks
- Timer
- CAP Stations

Air defence operators must collect as much information about incoming tracks as possible so that they are able to correctly identify and classify any aircraft.

NAUTICAL MILES

A nautical mile (NM) is equal to 1852 meters.

MARITIME TASK GROUP

A ship is part of a Maritime task group with may involve other ships in a close proximity. In this task you are operating for a ship in a maritime task Group

DATA LINK

When a possible track appears outside of the radar range it will appear as a data link.

The only information this data link will have is a bearing. You will be able to find out the bearing of the data link by sending out a sensor from the ship.

No other information can be found about the possible aircraft until the track enters the radar.

Ensure that the sensor and data link are correctly correlated based on the information provided.

TRACK

A track is an identified or unidentified aircraft which appears inside the radar screen.

FL

FL indicates the flight level. For example FL 250 is equal to 25,000 ft

AIR LANE

An airlane is a corridor in which general air traffic are routed by air traffic control.

In this task the airlane extends from FL 360 (36000ft) to FL 520 flying over the NFZ.

NO FLY ZONE

The NFZ operates over the border of country S and H for any aircrafts flying below FL 350 (35, 000 feet) in order to prevent aircraft injuring ground force and to engage in border conflict.

No aircraft should be flying below FL 350 in the vicinity of the NFZ regardless of whether they are in the airlane or not.

IFF

IFF stands for Identification Friendly or Foe. It is an identification system which enables operators to identify aircrafts and determine their bearing and range.

Aircrafts respond to radar interrogation with a message (squawk); military, friendly, or civil IFF.

Hostile aircrafts may not transmit any IFF.

IFF must correlate with the correct aircraft and flight profile. IFF in this task will be indicated by either unknown, friendly or hostile.

AIR INVESTIGATE PROCEDURE

An air investigate procedure is usually called out by the Principal warfare officer (PWO) or Air Warfare Officer (AWO).

When ‘Investigate’ is called out all members of the air defence operations team will try to accumulate as much information about the track as possible. This may include information about height, bearing, altitude, IFF etc.

CORRELATE

Tracks need to be correlated when two different sensors are indicating the same target.

Operators must be sure that the data provided from the sensors relates to just one track before they correlate the track.

CAP

CAP stands for a Combat Air Patrol which is an aircraft patrol provided over an air defence area for the purpose of intercepting and destroying hostile aircraft before they reach their target.

CAP should be made aware of all tracks flying in the vicinity of NFZ and any changes to altitude.

CAP’s are usually sent after all attempts to investigate and gather information has been made. In this task there are 4 CAP stations positioned in the NFZ. There are 2 CAP’s based at each CAP station.

CAP aircrafts may; **Shadow**, **Mark** or **Cover** the unidentified/hostile aircraft in question.

Shadow = Maintain (visual or radar) contact with specified target

Mark = An order to a missile equipped unit to initiate an engagement on a specified track up to but not including the point of firing

Cover = Directive to establish an air-to-air position that will allow engagement of a specified target or threat.

Break Off = Discontinue approach but maintain contact unless otherwise indicated.

RULES OF ENGAGEMENT

You must verify the ROE before you take action to engage in an aircraft.

FGA

A type of aircraft - Fighter Ground Aircraft

HOSTILE INTENT

The aircraft is displaying behaviours which lead you to believe the aircraft is hostile and its actions could lead to loss of life

AIR THREAT and SELF-DEFENCE WARNINGS

A ship’s company operates under three different ‘Air Defence Warning Levels’ (ADWLs). These are:

1. 'White' – Attack is unlikely without adequate warning

2. 'Yellow' – Attack is probable; and

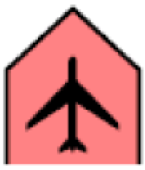
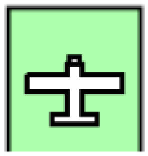
3. 'Red' – Attack on the ship is either imminent or is in progress.

Read warnings. Air threat warnings sent to aircrafts to attempt to establish their identity.

Self-defence warnings

This means that you have to give 'adequate warning' and only use force that is immediate, of overwhelming need and proportional

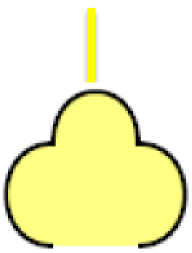
SYMBOLOGY

	UNKNOWN DATA LINK/TRACK	HOSTILE AIRCRAFT/FGA	Civial Aircraft/ CIVAIR	CAP stations
				

FLIGHT PROFILE

Each track has a velocity leader to give an indication of speed. The longer the line, the faster the track is moving.

Velocity Leader



TR: Track number

IFF: Unknown, Friendly or Hostile

FL: flight level e.g. FL150 = 15 000ft

Appendix 8b.1: Peace Enforcement scenario briefing

BACKGROUND:

There is a water supply which runs through both countries starting in Country *H* running down through farmland into Country *S*. A peace accord exists where H has agreed not to build any dams. Tensions have arisen between the two countries over water rights which threatens the accord.

Country H believes that they have traditional territorial water rights and have recently built a dam in order to prevent water from entering the southern country. The country is comprised of a Highland group that have resources and money and want to retain control of the water and maintain reputations.

Country S is occupied by a group of farmers who need the water for their crops and livestock. They are relatively peaceful people; however, tension has arisen over who has rights to the water which runs through their land.

Country S has known ties with other religious groups in the south. These religious groups have access to resources such as weapons and aircraft and are threatening to take action if the issue is not resolved.

There has been an appeal for assistance from the UN.

At present a NFZ has been put in place over the southern area of *Country H*

MISSION/TASK:

United Nations Security Council Resolution UNSCR have set a 'no fly zone' (NFZ) to all combat aircraft. You are tasked to monitor the activities from off the coastline of *Countries S* and *H* and to enforce the NFZ over the northern highlands of *Country H* from all combat aircraft entering the zone. The NFZ is within the range of missiles.

CURRENT SITUATION:

In recent months the leaders of *H* and *S* have made significant progress in key issues over water rights. A peace agreement has been signed which states that the Highland people will allow water to flow into *Country S*. However, 'a no fly zone' is still in place over *Country H.* Country *S* believes this to be unfair and strong tensions remain.

RULES of ENGAGEMENT

You may engage in the aircraft if you believe it is performing a hostile act or there has been a visual identification of an armed aircraft failing to respond to warnings and operating in a no fly zone and or you feel action is necessary for self-defence.

You will now take part in a series of practice trials.

There are 4 parts to this practice trial. You will be instructed when to turn over your answer sheet.

Do not do so until you see **"Please turn over Event ...."**

Please let the researcher know when you have finished reading this and take this time to ask any questions you may have.

Appendix 8c: Task Instructions – 20s/10s/attend/non attend

During this task you will be asked to take on the role of an air defence operator in a peace enforcement mission.

Your task will involve making a number of different decisions regarding the events that play out before you. We are also interested in your confidence in these decisions.

At certain time intervals the simulation will pause. During this time please turn over the question sheet with corresponding event number. The simulation will continue again after 5 seconds and you will be required to make a decision regarding what is presented to you on the screen and then rate how confident you are that you have chosen the best decision given the current situation. You have **20/10seconds** to make a decision and provide your confidence rating. You may have to make more than one decision during this time. You will be instructed on the screen if this is the case.

There will be a timer on the screen which will count down from **20/10s seconds**. Once the time is up the screen will blank out again. The scenario will then continue until the next decision point and you will repeat the process.

ATTEND: During the task you will wear a pair of headphones and listen to a conversation involving 2 pilots and air traffic control communications. **You must attend to the information you are hearing as there will be a series of questions that will be asked at the end of the scenario.**

DO NOT ATTEND: During the task you will wear a pair of headphones involving a conversation between 2 pilots and air traffic control communications. **This is background noise and it is not necessary for you to attend to it.**

The scenario has already been scripted and will run in a particular order regardless of the last decision you made. Therefore, please ensure that you are making decisions and rating your confidence at each decision point and do not reflect on previous decisions.

Appendix 8d: Audio attention questions for participants

1. At the start of the audio, was the air traffic controller male or female?

2. Did the air traffic controller changed during the scenario?

3. How long did the pilots keep on each fuel tank?

4. What events were going on in the area which pilots were told to be aware of?

Appendix 8e: Confidence-Accuracy Metacognitive Training

Instructions to participants (PowerPoint)

In this experiment you will be required to make a decision and then rate how confident you are in that decision. At each decision event you will be provided with three decision options and we would like you to choose the option you believe to be the best decision.

One of these options has been deemed the best response given the situation. Once you have made a decision we would like you to rate how confident you are that you have chosen the best decision.

The aim of this task is to see if the confidence rating you choose (perceived confidence) relates to the decision taken (actual accuracy) in a useful way (i.e., that you apply greater confidence to accurate decisions than those that are inaccurate).

The confidence rating is on a scale from 0 to 5. 0 being you are *'not at all confident' (*which should align more closely to *incorrect* decisions*)* and 5 being *you* are *'extremely confident'* (which should align more closely with what you feel are correct decisions).

Appendix 8f: Answer booklet for High TL condition

Event 1: The ship picks up information of a new data link outside the radar. The information is consistent with a civil aircraft following airlane.

Please place a tick next to the option you think would be best given your current situation

1. Examine the data link. ☐

2. Confirm that the track is following airlane. ☐

3. Conduct air investigate procedure by examining all the information from all sensors. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	2	3	4	5

Not at all confident   Extremely confident

Event 2: A sensor is sent from the ship on the bearing of the incoming track.

Please place a tick next to the option you think would be best given your current situation

1. Associate the sensor with link track. ☐

2. Correlate sensor with data link track. ☐

3. Analyse the sensor parameters for classification. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**

Not at all confident Extremely confident

Event 3: The sensor bearing remains consistent with the link track and paramaters consistent with civair aircraft

Please place a tick next to the option you think would be best given your current situation

1. Ignore new information. ☐

2. Correlate sensor information with background traffic. ☐

3. Correlate the sensor information with link track. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**

Not at all confident Extremely confident

Event 4: The track appears inside the ships' radar in the vicinity of the incoming link track already located by the sensors. The radar track is transmitting friendly identification friendly or foe (IFF)

Please place a tick next to the option you think would be best given your current situation

1. Ascertain contact altitude. ☐

2. Accept IFF identity. ☐

3. Do Nothing. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident			Extremely confident	

Event 5: A new data link appears on the screen

Please place a tick next to the option you think would be best given your current situation

1. Examine the information that is presented to you. ☐

2. Confirm track is following airlane. ☐

3. Conduct Air investigate procedure by examining all the information from all sensors. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**
Not at all confident Extremely confident

Event 6: Track 002 appears on long range radar. No altitude information

Please place a tick next to the option you think would be best given your current situation

1. Examine sensor contributions. ☐

2. Conduct an air investigate procedure by examining all the information from all sensors. ☐

3. Monitor track. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**

Not at all confident Extremely confident

Event 7: A new data link appears on the screen

Please place a tick next to the option you think would be best given your current situation

1. Examine the information that is presented to you. ☐

2. Confirm track is following airlane. ☐

3. Conduct Air investigate procedure by examining all the information from all sensors. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 8: Track 002 IFF information does not correlate with existing track.

Please place a tick next to the option you think would be best given your current situation

1. Conduct an air investigate procedure by examining all the information from all sensors. ☐

2. Call CAP to investigate. ☐

3. Engage. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 9: Track 003 appears on long range radar. No altitude information

Please place a tick next to the option you think would be best given your current situation

1. Examine sensor contributions. ☐

2. Conduct an air investigate procedure by examining all the information from all sensors. ☐

3. Monitor track. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 2 3 4 5

Not at all confident Extremely confident

Event 10: Unidentified track 002 reported at 14, 000 feet

Please place a tick next to the option you think would be best given your current situation

1. Engage. ☐

2. Send a fighter CAP to investigate. ☐

3. Attempt to establish communication and issue warnings. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 2 3 4 5

Not at all confident Extremely confident

Event 11: Track 001 continues in the airlane and starts to move over the NFZ

Please place a tick next to the option you think would be best given your current situation

1. Do nothing ☐

2. Advise combat aircraft patrol of assumed civair complying with airlane ☐

3. Continue to monitor ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**
Not at all confident Extremely confident

Event 12: Track 003 IFF information does not correlate with existing track.

Please place a tick next to the option you think would be best given your current situation

1. Conduct an air investigate procedure by examining all the information from all sensors. ☐

2. Call CAP to investigate. ☐

3. Engage. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**
Not at all confident Extremely confident

Event 13: Unidentified track 003 reported at 14, 000 feet

Please place a tick next to the option you think would be best given your current situation

1. Engage with the aircraft. ☐

2. Send a fighter CAP to investigate. ☐

3. Attempt to establish communication and issue warnings. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident			Extremely confident	

Event 14: Track 002 enters no fly zone

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to intercept and visually identify the aircraft ☐

2. Continue to Read Warnings. ☐

3. Order CAP to engage. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident			Extremely confident	

Event 15: As track 001 crosses over the national boarder it drops in altitude to 35, 400 ft.

Please place a tick next to the option you think would be best given your current situation

1. Continue to monitor and attempt to contact air traffic control. ☐

2. Do nothing. ☐

3. Advise combat air patrol of assumed civair complying with
airlane in change of flight level.

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not all confident			Extremely confident	

Event 16: CAP identifies aircrafts 002 as a fighter ground aircraft from Country S

Please place a tick next to the option you think would be best given your current situation

1. Order the fighter to verify weapon status.

2. Order CAP engagement.

3. Order fighter to attempt communication and read warnings.

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident			Extremely confident	

Event 17: Track 003 enters no fly zone

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to intercept and visually identify the aircraft. ☐

2. Continue to Read Warnings. ☐

3. Order CAP to engage. ☐

1	2	3	4	5
Not at all confident				Extremely confident

Event 18: Fighter reports that the aircraft 002 has weapons

Please place a tick next to the option you think would be best given your current situation

1. Verify ROE for engagement. ☐
2. Order CAP to establish an air to air position that will allow engagement. ☐
3. Order CAP to engage ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	2	3	4	5
Not at all confident				Extremely confident

Event 19: CAP identifies aircrafts 003 as a fighter ground aircraft from Country S

Please place a tick next to the option you think would be best given your current situation

1. Order the fighter to verify weapon status. ☐

2. Order CAP engagement. ☐

☐

3. Order fighter to attempt communication and read warnings.

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 20: Fighter reports that the aircraft 003 has weapons

Please place a tick next to the option you think would be best given your current situation

1. Verify ROE for engagement. ☐

2. Order CAP to establish an air to air position that will allow engagement. ☐

3. Order CAP to engage ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**

Not at all confident Extremely confident

Event 21: Aircraft 002 gradually alters course to port. No change in altitude. No communications

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to initiate an engagement on track 002 up to but not including the point of firing . ☐

2. Order CAP to engage. ☐

3. Order CAP to continue to read warnings. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**

Not at all confident Extremely confident

Event 22: Aircraft 003 gradually alters course towards the ship. No change in altitude. No communications

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to initiate an engagement on track 002 up to but not including the point of firing ☐

2. Order CAP to engage ☐

3. Order CAP to continue to read warnings. Warning 4 ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 23: Aircraft 003 steadies on a heading toward Maritime Task Group. No change in altitude

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to engage ☐

2. Order Air Threat Warning Red. Attack on ship is imminent ☐

3. Order Cap to initiate an engagement on a specified track up to but not including the point of firing and continue to read warnings ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 24: aircraft 002 leaves no fly zone heading south

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to continue to initiate an engagement on track 002 up to but not including the point of firing ☐

2. Order CAP to maintain (visual or radar) contact with specified targetfrom inside no fly zone ☐

☐

3. Order CAP to discontinue approach but maintain contact
unless otherwise indicated.

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**
Not at all confident Extremely confident

Event 25: FGA descends to low level and continues to close the task

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to engage ☐

2. Order Air Threat Warning Red. Attack on the ship is imminent ☐

3. Order Cap to Order Cap to initiate an engagement on a specified track up to but not including the point of firing and continue to read warnings ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1 **2** **3** **4** **5**
Not at all confident Extremely confident

Event 26: Aircraft 002 turns towards NFZ again

Please place a tick next to the option you think would be best given your current situation

1. Order CAP to read warnings. ☐

2. Engage ☐

3. Order CAP to establish an air-to-air position that will allow engagement of track 002 ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident			Extremely confident	

Event 27: FGA continues to descends to low level and continues to close the task group

Please place a tick next to the option you think would be best given your current situation

1. Engage with force weapons from the ship ☐

2. Cover with force weapons and Order CAP to discontinue approach but maintain contact unless otherwise indicated ☐

3. Read warnings and monitor for Hostile Intent . Set ship missiles on target. Self-defence warning ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 28: FGA continues to close the Maritime Task Group and demonstrates Hostile Intent

Please place a tick next to the option you think would be best given your current situation

1. Continue to read warnings ☐

2. Order CAP to engage. ☐

3. Verify that CAP is clear and Engage with Force weapons ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 29: A new data link appears on the screen

Please place a tick next to the option you think would be best given your current situation

1. Examine link track data ☐

2. Validate track is following airlane ☐

3. Conduct Air investigate procedure. ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Event 30: Track appears on long range radar. No altitude information

Please place a tick next to the option you think would be best given your current situation

1. Examine sensor contribution ☐

2. Conduct an air investigate procedure by examining all the information from all sensors ☐

3. Monitor track ☐

How confident are you that you have chosen the best decision given your current situation

Please circle

1	**2**	**3**	**4**	**5**
Not at all confident				Extremely confident

Appendix 9 (See Disc): 9a. High TL scenario Video, b. Moderate TL Scenario Video, c. Low TL scenario Video d. Practice Trial video.

Appendix 10a: Means and Standard Deviations for W-S C-A as influenced by DC, TL and group

TL	Group	Overall	High DC	Medium DC	Low DC
High	NG		.12 (.30)	-.01 (.27)	.19 (.20)
	VGP1		-.02 (.40)	-.13 (.41)	.02 (.43)
	VGP2		-.04 (.44)	-.01 (.31)	.03 (.30)
	Total		.02 (.38)	-.05 (.33)	.08 (.31)
Moderate	NG		.08 (.35)	-.06 (.36)	-.10 (.50)
	VGP1		.06 (.31)	-.05 (.35)	-.01 (.35)
	VGP2		.02 (.24)	-.06 (.33)	-.00 (.23)
	Total		.05 (.29)	-.06 (.33)	-.04 (.37)
Low	NG		-.03 (.33)	.07 (.21)	.14 (.41)
	VGP1		.11 (.30)	.00 (.31)	.06 (.26)
	VGP2		.10 (.29)	.21 (.21)	-.03 (.39)
	Total		.06 (.30)	.10 (.56)	.06 (.35)
	NG		.06 (.32)	.00 (.28)	.08 (.40)
	VGP1		.05 (.33)	-.06 (.35)	.02 (.34)

VGP2	.03 (.33)	.04 (.30)	-.00 (.30)
Total	.04 (.32)	-.00 (.31)	.03 (.35)

Note. Standard Deviations are in parenthesis

Appendix 10b: Means and Standard Deviations for dimensions of psychometric scores as influenced by group

Measure	NG	VGP1	VGP2	Total
Tolerance to Ambiguity	48.03 (7.48)	49.10 (6.96)	47.23 (6.91)	48.12 (7.08)
Decisiveness	22.67 (4.51)	22.50 (5.23)	23.67 (4.72)	22.94 (4.80)
Ambiguity Tolerance B	36.40 (6.03)	37.03 (7.58)	35.63 (6.95)	36.36 (6.83)
Decision style	59.17 (9.10)	59.50 (12.20)	59.30 (10.26)	59.32 (10.48)
Neuroticism	23.73 (8.31)	23.57 (10.44)	25.63 (10.80)	24.31 (9.84)
Extraversion	31.93 (6.09)	29.90 (6.96)	30.20 (5.66)	30.68 (6.26)
Openness To experience	31.60 (7.28)	33.43 (7.78)	35.37 (5.80)	33.47 (7.10)
Agreeableness	35.13 (5.20)	30.40 (6.42)	30.83 (6.25)	32.12 (6.29)
Conscientious ness	28.70 (8.11)	29.30 (8.73)	28.17 (7.09)	28.72 (7.93)

Note. Standard deviations are in parenthesis.

Appendix 10c: Means and Standard Deviations for W-S C-A as influenced by DC, TL, time and group

TL	Group	Time Condition	High DC	Medium DC	Low DC
High	NG	10s	.17 (.44)	-.10 (.33)	.01 (.26)
		20s	.12 (.30)	-.01 (.27)	.19 (.20)
		Total	.14 (.37)	-.05 (.30)	.10 (.25)
	VGP1	10s	.04 (.34)	-.09 (.35)	.00 (.46)
		20s	-.02 (.40)	-.13 (.41)	.02 (.43)
		Total	.01 (.36)	-.11 (.37)	.01 (.43)
	VGP2	10s	.01 (.36)	.01 (.31)	-.02 (.36)
		20s	-.04 (.44)	-.01 (.31)	.03 (.30)
		Total	-.01 (.39)	.00 (.30)	.00 (.32)
	Total	10s	.07 (.38)	-.06 (.32)	.00 (.36)
		20s	.02 (.38	-.05 (.33)	.08 (.32)
		Total	.05 (.38)	-.05 (.32)	.04 (.34)
Mod	NG	10s	.12 (.33)	.00 (.32)	.11 (.28)
		20s	.08 (.35)	-.10 (.32)	-.10 (.50)
		Total	.10 (.33)	-.05 (.31)	.01 (.41)

	VGP1	10s	.04 (.35)	-.09 (.37)	.04 (.36)
		20s	.06 (.25)	-.03 (.34)	-.01 (.35)
		Total	.05 (.30)	-.06 (.35)	.01 (.34)
	VGP2	10s	.20 (.21)	.09 (.31)	-.18 (.28)
		20s	.04 (.22)	-.04 (.32)	.00 (.23)
		Total	.12 (.23)	.03 (.31)	-.09 (.26)
	Total	10s	.12 (.30)	.00 (.33)	-.01 (.32)
		20s	.06 (.27)	-.06 (.32)	-.04 (.37)
		Total	.09 (.28)	-.03 (.32)	-.02 (.34)
Low	NG	10s	.19 (.36)	.17 (.27)	.03 (.28)
		20s	-.03 (.33)	.07 (.21)	.17 (.39)
		Total	.08 (.36)	.12 (.24)	.10 (.34)
	VGP1	10s	.18 (.34)	.08 (.31)	-.20 (.42)
		20s	.11 (.30)	.00 (.31)	.06 (.26)
		Total	.14 (.31)	.04 (.30)	-.07 (.36)
	VGP2	10s	.00 (.34)	.09 (.41)	.05 (.52)
		20s	.10 (.29)	.21 (.21)	-.03 (.39)

		Total	.05 (.31)	.15 (.32)	.01 (.45)
	Total	10s	.12 (.35)	.11 (.32)	-.04 (.42)
		20s	.06 (.30)	.10 (.26)	.06 (.35)
		Total	.09 (.32)	.10 (.29)	.01 (.39)
Total	NG	10s	.16 (.37)	.02 (.32)	.05 (.27)
		20s	.06 (.32)	-.01 (.27)	.09 (.40)
		Total	.11 (.35)	.01 (.29)	.07 (.34)
	VGP1	10s	.09 (.34)	-.04 (.34)	-.05 (.41)
		20s	.05 (.31)	-.05 (.35)	.02 (.34)
		Total	.07 (.33)	-.04 (.34)	-.02 (.38)
	VGP2	10s	.07 (.31)	.06 (.34)	-.05 (.39)
		20s	.04 (.33)	.05 (.30)	.00 (.30)
		Total	.05 (.32)	.06 (.31)	-.03 (.35)
	Total	10s	.11 (.34)	.02 (.33)	-.02 (.36)
		20s	.05 (.32)	.00 (.31)	.04 (.35)
		Total	.08 (.33)	.01 (.32)	.01 (.36)

Note. Standard Deviations are in parenthesis

Appendix 10d: Means and Standard Deviations for W-S C-A as influenced by to DC, TL, audio and group

TL	Group	Audio	Overall	High DC	Medium DC	Low DC
High	NG	Attend	.07 (.14)	.02 (.38)	.30 (.26)	-.08 (.43)
		Non Attend	.03 (.16)	.11 (.35)	-.07 (.40)	.18 (.32)
		Total	0.05 (.14)	.06 (.35)	.12 (.37)	.05 (.38)
	VGP1	Attend	.03 (.19)	.00 (.25)	.01 (.31)	.02 (.41)
		Non Attend	.03 (.23)	.06 (.47)	-.19 (.53)	.29 (.29)
		Total	.03 (.23)	0.03 (.36)	-.09 (.42)	.15 (.36)
	VGP2	Attend	-.01 (.22)	.26 (.40)	-.12 (.28)	-.08 (.17)
		Non Attend	.07 (.19)	.14 (.28)	-.01 (.31)	-.01 (.42)
		Total	.03 (.20)	.20 (.33)	-.07 (.28)	-.05 (.30)
	Total	Attend	.03 (.18)	.09 (.35)	.06 (.32)	-.05 (.33)
		Non Attend	.05 (.18)	.10 (.34)	-.09 (.40)	.15 (.35)
		Total	.04 (.18)	.10 (.34)	-.01 (.36)	.05 (.35)
Moderate	NG	Attend	.30 (.17)	.38 (.31)	.28 (.29)	.13 (.31)
		Non Attend	-.03 (.07)	.15 (.36)	.18 (.23)	-.16 (.20)
		Total	.17 (.21)	.29 (.33)	.24 (.26)	.02 (.30)
	VGP1	Attend	.05 (.22)	-.15 (.21)	.24 (.32)	.05 (.39)

		Non Attend	.11 (.19)	.18 (.38)	.22 (.34)	-.02 (29)
		Total	.08 (.17)	.01 (.34)	.23 (.31)	.02 (.22)
	VGP2	Attend	.01 (.26)	.19 (.19)	-.15 (.42)	-.01 (.39)
		Non Attend	0.10 (.20)	.20 (.41)	-.01 (.42)	.05 (.49)
		Total	.05 (.22)	.19 (.30)	-.08 (.40)	.02 (.42)
	Total	Attend	.13 (.24)	.15 (.33)	.14 (.38)	.06 (.29)
		Non Attend	.07 (.16)	.18 (.36)	.13 (.38)	-.03 (.34)
		Total	.10 (.20)	.16 (.33)	.13 (.35)	.02 (.31)
Low	NG	Attend	-.02 (.14)	.02 (.32)	.09 (.25)	.01 (.22)
		Non Attend	.19 (.19)	.26 (.15)	.09 (.38)	.30 (.33)
		Total	.08 (.19)	.14 (.26)	.09 (.30)	.16 (.31)
	VGP1	Attend	.11 (.13)	.21 (.33)	-.03 (.39)	.03 (.23)
		Non Attend	.21 (.16)	.17 (.43)	.13 (.40)	.19 (.28)
		Total	.16 (.15)	.19 (.36)	.05 (.38)	.11 (.26)
	VGP2	Attend	.04 (.16)	-.01 (.06)	.11 (.48)	.00 (.30)
		Non Attend	.15 (.12)	.30 (.26)	.17 (.21)	-.06 (.39)
		Total	.11 (.14)	.18 (.25)	.14 (.32)	-.04 (34)
	Total	Attend	.04 (.14)	.08 (.27)	.05 (.35)	.02 (.23)

	Non Attend	.18 (.15)	.24 (.28)	.13 (.31)	.13 (.36)
	Total	.12 (.16)	.17 (.29)	.09 (.33)	.08 (.31)
Total	Attend	.13 (.20)	.15 (.36)	.23 (.27)	.03 (.32)
	Non Attend	.07 (.17)	.17 (.28)	.06 (.34)	.13 (.34)
	Total	.10 (.19)	.16 (.32)	.15 (.31)	.07 (.33)
VGP1	Attend	.06 (.16)	.02 (.29)	.07 (.34)	.03 (.27)
	Non Attend	.12 (.19)	.14 (.40)	.05 (44)	.15 (.30)
	Total	.09 (.18)	.08 (.35)	.06 (.39)	.09 (.28)
VGP2	Attend	.01 (.20)	.16 (.27)	-.07 (.38)	-.03 (.28)
	Non Attend	.11 (.16)	.22 (.31)	.06 (.31)	-.01 (.40)
	Total	.06 (.19)	.19 (.29)	-.00 (.34)	-.02 (.35)
Total	Attend	.07 (.19)	.11 (.31)	.08 (.34)	.01 (.29)
	Non Attend	.10 (.17)	.18 (.33)	.06 (.36)	.09 (.35)
	Total	.09 (.18)	.14 (.32)	.30 (.36)	.05 (.31)

Note. Standard deviations are in parenthesis

Appendix 10e: Means and Standard Deviations for dimensions of SA as influenced by MFT and group

MFT	Group	Demand	Supply	Understanding
Yes	VGP	12.67 (3.13)	18.73 (2.74)	12.73 (3.71)
	NG	13.53 (3.48)	19.60 (3.64)	11.60 (3.09)
	Total	13.10 (3.28)	19.17 (3.20)	12.17 (3.41)
No	VGP	12.00 (4.31)	18.27 (3.47)	11.33 (3.54)
	NG	14.27 (4.06)	19.00 (3.16)	12.20 (3.21)
	Total	13.13 (4.27)	18.63 (3.29)	11.77 (3.35)
Total	VGP	12.33 (3.72)	18.50 (3.08)	12.03 (3.63)
	NG	13.90 (3.74)	19.30 (3.37)	11.90 (3.11)
	Total	13.12 (3.78)	18.90 (3.22)	11.97 (3.36)

Note. Standard Deviations are in parenthesis

Appendix 11: Dstl Report Documentation

[Report classification]

Report documentation page v5.0

* Denotes a mandatory field

1a.	**Report number: *** N/A	**1b.**	**Version number:** 1
2	**Date of publication: *** 03/01/2019	**3.**	**Number of pages:** 367
4a.	**Report UK protective marking: ***	UK OFFICIAL	
4b.	**Report national caveats: ***	NONE	
4c.	**Report descriptor: ***	NONE	
5a.	**Title: *** PhD Thesis: Measuring the Decision Accuracy and Decision Confidence of Air Defence Operators		
5b.	**Title UK protective marking: ***	UK OFFICIAL	
5c.	**Title national caveats: ***	NONE	
5d.	**Title descriptor: ***	NONE	
6.	**Authors: *** J.E Adams-White		
7a.	**Abstract: *** Air defence decision making is demanding, time pressured and complex. Operators must make complex and cognitively demanding decisions in dynamic and uncertain environments. This thesis sought to increase the understanding of, and measure, decision confidence and accuracy in air defence decision making. In doing so, a novel method was designed and developed which was based on an integration of Classical Decision Making (CDM) and Naturalistic Decision Making (NDM) theories of decision making. The experimental stimulus was developed with the help of Subject Matter Experts. This included the development of a realistic computer-generated air defence scenario in which the Task Load (high, moderate, low) was varied. In addition, an appropriate range of classifiable decision options, which varied in Decision Criticality (high, medium, low) was also generated. Three-hundred novice participants and twenty-two experts took part across a range of experiments. Recommendations suggest the findings can be applied to training and decision support technology. Further, the outcomes generally support the potential use of more traditional experimental methods alongside naturalistic approaches in critical environments, and that such an approach is warranted.		
7b.	**Abstract UK protective marking: ***	UK OFFICIAL	
7c.	**Abstract national caveats: ***	NONE	
7d.	**Abstract descriptor: ***	NONE	
8.	**Keywords:** Metacognition, Decision Making, Air Defence, Confidence, Training,		

Please note: Unclassified, Restricted and Confidential markings can only be used where the report is prepared on behalf of an international defence organisation and the appropriate prefix (e.g. NATO) included in the marking.

* Denotes a mandatory field

9.	**Name and address of publisher: *** The University of Liverpool School of Engineering Brownlow Hill Liverpool L69 3GH	**10.**	**Name and address of funding source:** Defence Science Technology Laboratory Porton Down, Salisbury SP4 0JQ
11.	**Funding source contract:**	MOD CSA S&T	
12.	**Dstl project number:**	706013	
13.	**Programme:**	Above Water	
14.	**Other report numbers:**	None	
15a.	**Contract start date:** 02.10.2014	**15b.**	**Contract end date:** 03.01.2019
16.	**IP conditions for report: ***	DEFCON 705 Full Rights	
17a.	**Patents:**	No	
17b.	**Application number:**	Not applicable	
18.	**Release authority role:**	University of Liverpool	