The Air Defence Task: Understanding what motivates automation usage to support classification decisions in practice

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy
by
Chloë Barrett-Pink

June 2018

Abstract

With the growing use of automated systems in military environments it remains vital that research continues to explore the use of such systems operationally. The recent literature has tended to take a systems focused approach, which has concentrated on features of the system and what impact alterations have upon task performance. However, research has begun to see the value in taking a human-centred perspective to understanding the use of automated systems in military environments; asking research questions that remain focused on the human operators that are required to utilise automated systems in increasingly complex environments. Therefore, this thesis contributes to the literature on human-machine-interaction through exploring the operational use of automated systems in the maritime environment. Research into sociotechnical systems is complex, therefore this thesis adopted a Naturalistic Decision Making (NDM) approach utilising mixed-methods to elicit understanding and knowledge from unique access to Royal Navy (RN) Subject Matter Experts (SME). Privileged access to a large number of RN experts (N=53) enabled novel and interesting findings to be drawn from two qualitative surveys. The first explored the stages of the air defence task conducted by RN personnel to better understand where uptake of automation may be beneficial. The findings of this questionnaire revealed that the high-level stages of the air defence task (Observe, Identify and Classify, and Decide and Act) have remained unchanged over the last 20 years and the areas that have previously been identified as potentially benefiting from automated system support remain the same. These findings raised pertinent questions as to why the same areas are still in need of support. Therefore, the second study of this thesis aimed to explore where automated systems have been brought into service to support RN operations to understand how the current procurement process functions. A second questionnaire was developed which allowed RN SME to discuss how automated systems are currently used across all operational settings and where they may be used in the future. Crucially, this second questionnaire explored RN SME opinions towards the existing procurement process. Of concern was that the findings of this study revealed the disconnect that often exists between end user and system designer which has a negative effect on the development of systems being fit for purpose at time of release. This in turn can have severe negative consequences to capability, appropriate system use and can increase the financial costs of developing and implementing new systems. The findings from the first two studies presented in this thesis highlighted the need for, and recommended an increase in use of, immersive simulation environments to support automated system development and research. Therefore, the third part of this thesis presents the development and validation of a simulated microworld, the Automatic Radar Classification Simulation (ARCS). ARCS was designed by the author of this thesis in collaboration with a software engineer to replicate aspects of the air defence task conducted by RN personnel. This design process included 2 pilot studies, the results of which informed developments and changes to ARCS. Overall, this design process took 8 months with several iterations of ARCS being developed. Following the development stage, an experiment was conducted (N=42 university students) to validate the utility of ARCS as a microworld using a holistic real-time scenario to explore individuals rationales for using a generic automated system when performing a threat detection task. In line with previous research, participants cited workload and managing uncertainty as reasons for selecting to use the automated decision support system. However, unexpectedly task performance was not significantly improved with access to the support system and strong learning effects were observed. Overall, this thesis supports the newly proposed move away from traditional "levels of automation" approaches, advocating for taking a more holistic approach to research into human-machine-interaction. This can be achieved through promoting long-term and continuous engagement between end users and system designers, ensuring that a human-human relationship is maintained throughout the life-cycle of the automated system. Additionally, this thesis highlights the importance of effective communication within and between the military, industry and academia, and the negative implications that ineffective communication has upon naval capability. Finally, this thesis supports the literature that highlights the importance of training in immersive environments and has provided academia with a high-fidelity microworld with which to explore operator use of automated decision support systems in the maritime environment.

Contents

List of Figures

List of Tables

Acronyms

ADT	Air Defence Task
ARCS	Automatic Radar Classification Simulation
AUD	Automation Usage Decisions
BIS	Barrett Impulsivity Scale
C2	Command and Control
CDM	Critical Decision Method
CF	Cognitive Flexibility
CMS	Combat Management System
DSS	Decision Support System
FOST	Flag Officer Sea Training
LOA	Level of Automation
MA	Military Advisor
MOD	Ministry of Defence
MWC	Maritime Warfare Center
MWS	Maritime Warfare School
NDM	Naturalistic Decision Making
NFC	Need for closure
NM	Nautical Miles
RAP	Recognised Air Picture
RM	Royal Marines
RN	Royal Navy
ROE	Rules of Engagement
RPD	Recognition Primed Decision making
SA	Situation Awareness
SME	Subject Matter Experts
TADMUS	Tactical Decision Making Under Stress
TE	Threat Evaluation
TEWA	Threat Evaluation and Weapon Allocation
WA	Weapons Allocation

Acknowledgement

I would first like to thank my supervisors Prof. Laurence Alison and Prof. Simon Maskell, without whom this thesis would not have been possible. Laurence, thank you for your guidance and support over these last 4 years. You have provided me with the opportunity to grow and develop as a researcher, not just with work relating to this thesis topic but in all the interesting research at CAMI. Simon, you have given me insight into the workings of engineers and computer scientists, quite often donning the hat of a translator. Without your ability to articulate complex concepts in a way that can be understood across academic domains, my transition from psychology masters student to interdisciplinary researcher would have been far more challenging.

A special thank you to all the personnel at the Maritime Warfare Centre who facilitated this work and to the operators who took part in my studies. It has been a privilege to work with you all. To Dstl, whose funding made this project a reality, thank you for the opportunity to undertake this research.

Thank you to Tony McCabe who made ARCS a reality, you took my sketches and ideas and made them into a tangible microworld. Thank you also to the research groups across the EEE and psychology departments. A special thanks to Grace Carter, Lauren Swan and Neil Shortland for being constant sounding boards. Thank you also to Christine Unterhitzenberger, as housemates and now friends, you are ever my role model in organisation. Your advice of writing 400 words a day is what has got this thesis finished.

To my family and friends who are always there to provide a break from the world of academia. Your unwavering support has kept me on this path and I am eternally grateful. And last but by no means least, thank you to my partner in all things John. You have provided constant support and counsel along this roller coaster of a journey, alongside new jobs and getting on the property ladder, and thank you for the consistent supply of tea!

Publications

i **Barrett-Pink, C.,** Alison, L., & Maskell, S. (2017). The Air Defence Task: Understanding the cognitions that underpin automation usage to support classification decisions in practice. *Proceedings of the 13th International Conference on Naturalistic Decision Making.* Bath.

ii Waring, S., Alison, L., Carter, G., **Barrett-Pink, C.,** Humann, M., Swan, L., & Zilinsky, T. (2018). Information sharing in interteam responses to disaster. *Journal of Occupational and Organizational Psychology,* DOI: https//doi.org/10.1111/joop.12217

iii Shortland, N., Alison, L., & **Barrett-Pink, C.** (2018). Military (in)decision-making process: a psychological framework to examine decision inertia in military operations. *Theoretical issues in Ergonomics Science, 19*(6), 752-772. DOI: 10.1080/1463922X.2018.1498826

iv **Barrett-Pink, C.,** Christiansen, P., Alison, E., & Alison, L. (under review). An Evidence based study of the positive impact of ORBIT and UK Polices Alcyone Training Program on Counter Terrorism Interviewing.

v **Barrett-Pink, C.,** Alison, L., & Maskell, S. (2018). On the bridges - insight into the current and future use of automated systems as seen by Royal Navy personnel. Manuscript in preparation.

Chapter 1

When above water warfare meets human-machine-interaction: an overview of what is known and what remains to be answered

1.1 Introduction

The focus of this thesis is in exploring the operational application of automated systems in the military maritime above water environment. This thesis aims to firstly describe the environment naval personnel are operating in by investigating the current decision stages that occur when personnel conduct the air defence task. Secondly, this thesis will explore and discuss the current state of automated systems and their potential future state. Uniquely this thesis will explore Royal Navy (RN) personnel's experiences of the current procurement process. Finally, in response to the findings of the first two aims, this thesis will present the development of an above water maritime micro-world that will support academia in understanding how individuals interact with decision support systems in an ecologically valid and holistic way. This introductory chapter will outline the problem space that above water maritime operations sits in before presenting the psychological literature that has informed this thesis.

1.2 Problem Space

The remit of the UKs Royal Navy is vast, encompassing many roles and duties *"characterised by rapid, unpredictable changes and increasing global interdependence"* (Ministry of Defence, 2011, p. 1-1). However, battle-spaces are becoming less clearly defined (Ministry of Defence, 2014) compared to historical warfare and recently there has been an increase in joint operations. For example, CTF 150 a joint task force which currently operates over two million square miles around the Indian Ocean. This task forces' primary role is to counter terrorism and prevent trafficking of drugs and weapons that fund terrorist organisations (Royal Navy., nd). Conflict is increasingly dynamic and challenging; and requires operations and actors to adaptively adjust to context, circumstances and team structure (Hutton et al., 2017; Mendonça, 2007), *"Teams will need to be 'expert teams' that are flexible and able to cope with change, both predicted and unpredicted"* (Brown, 2007, p. 379). Adaptation has become a fundamental requirement to ensuring mission success; adaptable expertise from actors and adaptable automated systems.

In October 2016, a UAE vessel was hit by a guided missile fired from a Yemeni rebel boat, injuring one crew member (Binnie, 2017). Shortly following this incident, USS Mason deployed countermeasures in response to suspected anti-ship missiles fired from the Yemini shore by Houthi rebels (Huffington Post, 2016). These incidents serve to highlight *"that the anti-ship missile threat is increasingly transcending definitions of conventional and asymmetric warfare"* (Scott, 2017). Each naval vessel performs a specific role within the fleet. Several different ship classes will be mentioned in this thesis and Table 1.1 provides an overview of the roles of each ship class discussed.

Within recent history there have been a number of cases that have highlighted the fatal consequences when errors transpire in maritime operations. One such incident occurred on 17th May 1987 when the USS Stark was hit by two Exocet missiles fired from an Iraqi Mirage F1 (Brummer and Hirst, 1987; LaGrone, 2017). The USS Stark was operating 20 Nautical Miles (NM) outside of the Iranian declared war zone during the Iran-Iraq war when at around 21:09 the first Exocet hit the ship, 30 seconds later a second Exocet hit the target. The formal investigation into the incident revealed that miscommunication, non-consistent monitoring of the aircraft and errors in operating procedures resulted in the Stark being hit without taking any evasive manoeuvres or defensive actions - 37 crew lost their lives that day (Department of Defense, 1987). The full declassification of the report into the HMS Sheffield incident during the Falklands campaign echoes these findings. Members of the crew were disengaged from their tasks, being bored and a little frustrated by inactivity, leading to the ship being unprepared for an attack; no evasive manoeuvres or defensive actions were taken which resulted in the loss of 20 RN personnel (Cobain, 2017).

1.2.1 Current and Future Capabilities

"Advantages arise from the location, from the organization and from having either greater or better forces [1]"

Radar use is widespread as it is an effective way to detect the presence of objects at range where visual detection is either difficult or impossible. However, radar is inher-

[1](Machiavelli, 2006) p97.

Table 1.1: Types of ship classes that are mentioned within this thesis

Ship Class	**Role of Ship**	**Examples**
Destroyer	Destroyers are fast and manoeuvrable warships that provide air defence protection to larger vessels in a fleet or to escorted vessels.	Examples from the Royal Navy include: • Country-class Destroyers which operated between 1962 to 2006, including *HMS Glamorgan*. • Type 42 Destroyers which operated between 1975 and 2013, including *HMS Sheffield*. • Type 45 Destroyers (Daring Class) which have been operating since 2010 and includes *HMS Daring, HMS Dauntless, HMS Diamond, HMS Dragon, HMS Defender, HMS Duncan.* • An example from the United States Navy is the *USS Mason* which has been operating since 2003.
Frigate	In modern Navies Frigates tend to perform the role of an escort vehicle. Providing protection to other military vessels or merchant ships.	An example from the United States Navy is the *USS Stark* which operated from 1982 to 1999.
Cruiser	A type of warship that is similar to a Destroyer class and is used to provide air defence cover to the fleet. The USA and Russia are the only navies that continue to operate cruisers.	An example from the United States Navy is the *USS Vincennes* which operated between 1985 and 2005.
Aircraft Carrier	The role of an aircraft carrier is to project air power by providing Naval forces with a seagoing airbase.	The Royal Navy has had several classes of aircraft carrier including: • Invincible class carriers which operated between 1980 to 2014 included *HMS Invincible*. • In service now are the Queen Elizabeth-class carriers, commissioned in 2017 for example *HMS Queen Elizabeth*.
Mine countermeasures vessels	This type of ship is used to minesweep and safely remove and/or neutralise active underwater mines.	An example from the Royal Navy is *HMS Cattistock*.

ently uncertain. Signals are transmitted from the radar towards an object, part of that signal is reflected to be received and interpreted. However, signals have an associated measurement error typically based upon the beam width. If an aircraft is flying below the radar line, which is tangent to the surface of the Earth, it may not be detected due to the effects of the radar horizon. It is also possible for reflected signals from the surface to produce a radar return that can be misinterpreted as multiple targets. Although uncertain, radar remains the primary sensor and data input used to develop maritime situational awareness at range. Within commercial Air Traffic Control recent evidence suggests that the accuracy of Automatic Dependent Surveillance-Broadcast (ADS-B) data could replace current radar operations, however, caution should be applied as Borst et al. (2017) point out, *"studies in Europe and Asia have reported frequently occurring ADS-B position errors reaching up to 7.5 NM"* (p5). Although sensor technology continues to improve, there remains uncertainty with the fusing of data relating to air track movements.

Several examples from the Falklands war (2nd April 1982 - 14th June 1982) highlight the complexities associated with maritime defence. For example, HMS Invincible fired a Mark 46 torpedo and depth charge at what was thought to be a submarine periscope but, with hindsight, was a whale (Inskip, 2002). Anecdotally, instances of sonar errors leading to torpedo strikes against whales and radar errors resulting in missile firings at flocks of birds have occurred on multiple occasions. These false alarms and erroneous firings are due to a combination of factors, including the heightened stress levels during a conflict situation as well as due to sensor uncertainty. For example, radar returns can be received from a wide range of objects, including hostile actors but also weather patterns or animals. This can negatively impact operator situational awareness as they are incorporating uncertain information into the operational picture. This can have a dramatic impact on the crew and immediate future operations of that vessel. For instance, it could potentially cause a fratricide incident or result in the loss of ammunition that may be essential if the ship is required to defend itself against further threats. Reduced ammunition will limit the ships options for avoiding and/or dealing with future threats. Heightened vigilance during conflict situations combined with radar uncertainty can increase the number of false alarms which can have a negative knock on effect to further contact identification. This sadly was the case during the Falklands conflict. In an account, several years following the end of this conflict, the Ministry of Defence (MOD) reported that HMS Invincible had sightings of two incoming Exocet missiles 19 minutes before HMS Sheffield was hit, resulting in the tragic loss of 20 lives and injuries to 24 other crew on-board (Ezard, 2000). The number of fatalities and injuries was partly due to HMS Sheffield operating without a fully operational radar and the crew being on half-alert; they had less than a minute of warning before being hit. The warning that should have come from HMS Invincible did not arrive due to Command attributing the radar tracks to being spurious. Had that warning arrived, HMS Sheffield would have had up to 19 minutes to prepare their defensive posture, and may have been able to deploy countermeasures and manoeuvre to reduce the likelihood of being hit. Ships have three countermeasure options that can be deployed in an emergency situation to cause confusion, distraction or seduction. Confusion involves deploying chaff to provide a number of realistic false targets to prevent or delay acquisition of the real target. Distraction can be effective against missiles during the target search phase, by surrounding the vessel with decoys that are hopefully locked onto by the missile. Finally, seduction is designed for when a missile is locked onto the vessel. A large decoy is propelled from the vessel in an attempt to lure the missile to follow the decoy. Each countermeasure has maximum effectiveness if used at the optimum time. For example, distraction measures can only be deployed before the missile is locked onto the target vessel. Therefore, the criticality of time in conflict situations cannot be emphasised enough.

In his book on his experiences during the Falklands War on board HMS Glamorgan, Inskip (2002) writes

> "I noticed the faintest of small blips [radar detection], about the size of echo given by an albatross. It was unusual, albeit not unknown, to detect birds at that range ... At the next sweep of the radar it was gone and I breathed a sigh of relief but this was shattered one sweep later when, on the same baring and closer in, a firm echo painted for the first time. In my heart of hearts I knew it was an Exocet." (p157).

Shortly after this initial identification, HMS Glamorgan was hit by an Exocet before she was able to complete evasive manoeuvres - 14 crew members were killed. It should be noted that Inskip refers to the raw radar data that used to be analysed by officers on board warships. Current radar data is digitally processed, removing the requirement for current RN personnel to read raw data.

Instances of false positive judgements have also sadly occurred, for example during the Iran-Iraq War on the 3rd July 1988 the USS Vincennes incorrectly considered an Iran Air Airbus A300 to be a threat which they shot down, killing 290 people over the Arabian Gulf (Hammond, 2017). These cases highlight the difficult position operators are faced with and the consequences to themselves (if false negative decisions are made) or to civilians (if false positive decisions are made).

To add to this complex environment are the progressively sophisticated threats that are operating within the air defence domain. Current supersonic missiles, for example

Yakhont/Onyx/BrahMos, are capable of reaching up to Mach 3 (1029 m/s) operating at ranges of up to 300 km [2]. If a radar system is able to detect potential threats up to 400 km away, the operations team would have just over 4 minutes to detect the threat, identify as hostile, perform threat evaluation, allocate and launch the appropriate hard or soft kill countermeasures (see Table 1.2). However, to defeat the long-range detection of radars, threats can use the curvature of the earth to cruise below the radar horizon[3], preventing detection until much later in an engagement. For example, if a threat (i.e. missile) was cruising at 15 ft and the detecting radar stood at a height of 123.3 ft, with the missile travelling at Mach 3, the ship would have around 33 seconds from radar detection to the missile hitting the ship. (see Figure 1.1).

Table 1.2: Types of countermeasure used in maritime operations

Hard kill countermeasure	an effect that physically counters an incoming threat. For example, a missile that is fired to destroy an incoming missile.
Soft kill countermeasure	an effect that aims to cause the incoming threat to be confused, distracted or seduced away.

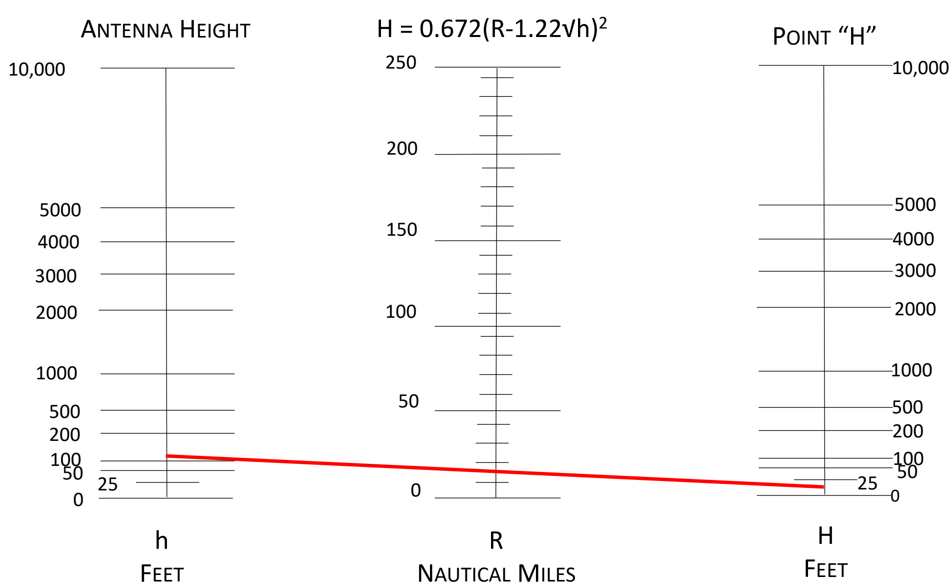


Figure 1.1: Earth Curvature Nomograph (adapted from www.rfcafe.com). The red line running from the height of the radar (123.3 ft) to the height of the target, in this instance the missile cruising at a height of 15 ft, indicates that the radar could detect the threat at 18.4 NM. If the missile was travelling at Mach 3, this would mean 33 seconds from detection to missile hit.

[2]https://www.janes.com/article/45334/brahmos-tested-to-the-limit-id14d1

[3]$R_{NM} = 1.23(\sqrt{hradar} + \sqrt{rtarget})$

Future missile systems currently under development are aiming to reach Mach 5-7 (1715 m/s-2401 m/s) (McDermott, 2017), detection at 400 km would give the operations team 2-3 minutes to respond and counter each threat. This time is further reduced if these threats are capable of operating below the radar horizon, severely limiting the ships ability to respond. For example, if we increase the speed at which the missile was cruising at 15 ft in the previous example to Mach 5, the ship would have 19.8 seconds from detection to missile hit. The loss of 37 lives aboard USS Stark during the Iran-Iraq war is testament to the danger of threats operating at sea-skimming levels. Since the USS Stark incident the American Navy has invested heavily in increasing defensive capability against standard anti-ship missile threats. However, recently Russia announced the successful testing of the Zircon (3M22 Tsirkon) hypersonic missile claiming to have reached speeds of up to Mach 8 (McDermott, 2017). This class of missile pose an equally significant threat due to their speed which severely limits operator reaction time. Correlational analysis of radar returns increases in uncertainty due to the vast distances that can be covered by the missile between radar sweeps. To counter this danger, the requirement to detect or identify threats at greater distances from the ship is crucial.

It is not only air defence that ships have to contend with. Increasingly close-range surface threats are highlighting potential areas of weakness in the overall defence of a ship or fleet. The diverse threats crews face requires a complete and constant understanding of the current environment. Due to the inherent uncertainty with radar and the speed of current and future threats, automated systems are essential to enabling complete situational awareness to be formed and maintained by the crew. New automated systems and robotics are transforming the art of war (de Boisboissel, 2017).

A multitude of actors may be in play at any one time which requires personnel to be continually updating their understanding of the environment, mission parameters and possible courses of action. The development of Situation Awareness (SA), defined as *"the perception of the elements in the environment within a volume of time and space, the comprehension of meaning and the projection of their status in the near future"* (Endsley, 1988) is crucial to facilitating effective decision making (Figure 1.2).

The research that Endsley and colleagues have conducted over the last 25 years has highlighted the importance of SA in facilitating the performance of military personnel and the achievement of their overall mission objectives (Endsley, 1995, 2015c). In fact, Endsley herself has highlighted how the terminology arose from military personnel (Endsley, 2015a) and remains commonly used, thereby supporting the concepts validity. This body of research has also identified several challenges associated with developing

SA including, data overload and automation design and implementation, both of which will be discussed in greater detail below. SA is intrinsically linked to decision making, *"decisions are formed by SA and SA is formed by decisions."* (Endsley, 2000, p. 5) however, it remains useful to view SA and decision making as individual concepts. For instance, it is possible to have a thorough SA of the environment and still make an incorrect decision, and vice versa (although less common) to make the correct decision and hold inaccurate or incomplete SA. Theoretically, SA and decision-making influence upon each other in a cyclical manner, yet can be decoupled through the impact of various factors. Chapter 3 discusses further the complexities associated with operating in the air defence theatre, the stages of decision making that exist and the facilitators and barriers to effective decision making at each stage. The next section of this chapter presents a hypothetical scenario to put the information presented so far into context.

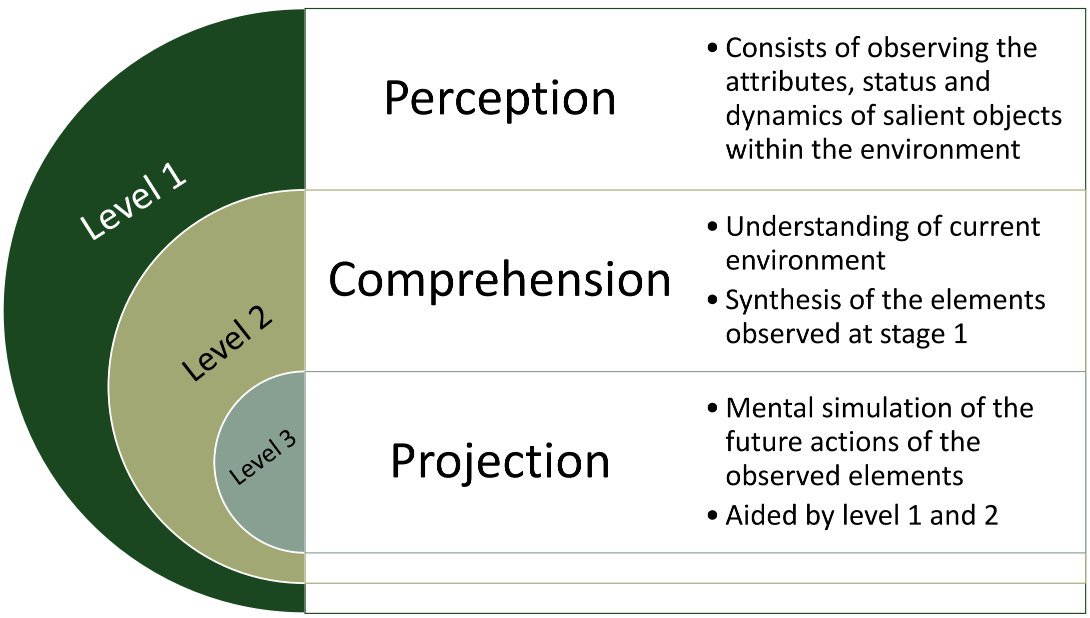


Figure 1.2: Levels of Situation Awareness (adapted from Endsley, 1988)

1.2.2 The Importance of Synergy between Man and Machine

Imagine a Type 45 Destroyer (T45) is operating in a littoral environment - an area of water close to land. The operations team are monitoring the incoming information on the Combat Management System (CMS). There are two major airports operating in the vicinity, with an average of 1250 flights a day [4]. The team have been operating in this vicinity for a week and there has been no suspicious or hostile activity. It is now the daily handover: one team has been operating through the morning and the afternoon team are about to take over control. Half an hour into the afternoon teams duties,

[4]http://www.express.co.uk/news/uk/475537/London-Heathrow-the-best-facts-stats-and-trivia-behind-the-UK-s-busiest-airport

and whilst they are finalising their understanding of the current environment, multiple suspicious aircraft (from this point on referred to as tracks) are picked up by the sensors. The operators monitoring the radar picture identify two tracks flying in formation on a direct heading towards the T45. The tracks are currently 50 NM from the ship and travelling upwards of 450 knots. This behaviour would be typically associated with military action, i.e. two military aircraft flying in attack formation. As the team are monitoring these tracks and continue to scan the environment, they identify three further potentially suspicious tracks travelling along the coast at low altitude. This is typical behaviour for a military aircraft that is attempting to use clutter from the land to mask its approach. They continue to monitor the five potentially hostile tracks observing standard Rules of Engagement (ROE). The initial two tracks increase their speed and descend to 500 ft, as they reach 15 NM from the T45 they turn away and increase their altitude, behaviour that would symbolise a release of weapons. Simultaneously a salvo of incoming missiles is picked up by the radar at close range. Whilst the operators were monitoring the suspicious tracks initially identified, they failed to pick up a sixth track that had been travelling within a designated airlane. This track had already fired a salvo of supersonic high dive missiles before turning away. The team have now just seconds to try and neutralise the imminent threat. The requirement for the appropriate use of automated systems to aid air defence operations is not just helpful, it is essential to increase the probability of survival.

As the example highlights, time pressures are critical. The complexity of the environment is caused not only by the number of actors that can be in play and the size of operational areas, but by the inherent uncertainty of radar and sonar data. The speed at which a potentially hostile situation can become critical places the operation teams under insurmountable cognitive load. Fighting power is built on capability (Ministry of Defence, 2014), a capability that can only be extended through the synergy of man and machine. Therefore, this thesis, in line with research exploring human machine interaction, argues that understanding when and why individuals would use an automated support system can further our understanding of how better to integrate and deploy such tools.

1.2.3 Historical research into use of automated systems

The area of Human Factors research is considered to have originated during the second world war (Wickens and Hollands, 2000) and focuses on exploring the relationship between humans and systems in sociotechnical environments. How interfaces are designed, what systems are developed and how they are updated, maintained and improved, are all part of this body of research. Research into human-machine-interaction remains topical due to the proliferation of automated systems across all aspects of society (Roth and Pritchett, 2017). Specific to military use of automated systems one focus is on facilitating the processing of the sheer volume of data that must be managed in order to build and maintain SA and increase capability to ensure operational success (Hawley et al., 2005; Ministry of Defence, 2011; Nguyen, 2002).

However, research conducted into human-machine-interaction argues that instead of automation being presented as *"a technology that executes a function that was previously performed by humans"* (Parasuraman and Riley, 1997), automated tools and systems change task structure. Arguably such systems do, in certain environments, replace the human, moving the human from the role of operator to one of being a supervisor to the system (Ghazizadeh et al., 2012). It has therefore been argued that it is important to conceive human-machine-interaction as an integrated series of tasks and subtasks allocated between the team dyad (Hoc, 2000). One way in which the concept of tasks and subtasks has been described and translated from the human-machine-interaction community to wider audiences is through the categorisation of levels of automation.

Table 1.3: Taxonomy of Automated Decision Action and Selection from (Parasuraman et al., 2000)

Low Automation	1	The computer offers no assistance: human must take all the decisions and actions.
	2	The computer offers a complete set of decision/action alternatives, or
	3	Narrows the selection down to a few or,
	4	Suggests one alternative,
	5	Executes the suggestion if the human approves, or
	6	Allows the human a restricted time to veto before automatic execution, or
	7	Executes automatically, then necessarily informs the human, and
	8	Informs the human only if asked or
	9	Informs the human only, if it, the computer, decides to
High Automation	10	The computer decides everything, acts autonomously, ignoring the human.

Currently there is a drive to increase the Level of Automation (LOA) systems have, which is posited to increase the support that systems can provide to the operator. Table 1.3 details the most commonly cited LOA proposed by Parasuraman et al. (2000). These levels highlight how automation function acts on a continuum, from no automation (1 on the scale) to complete automation (10 on the scale). However, recent concerns have been raised with regards to the predictive power of LOA based predictions of humans interaction with automation, and therefore the utility of taking such an approach to system design (Jamieson and Skraaning, 2017). LOA assumes linear and hierarchical concepts of how automation functions alongside personnel operating in the field, this is however not a true reflection on how systems are used operationally. Research has found that automated systems tend not to be employed at certain task levels but through operator experience features of the system are employed in an intuitive and fluid manner (Abbott, McKenney, and Railsback, 2013, as cited in Jamieson and Skraaning 2017). This mirrors the findings from Naturalistic Decision Making (NDM) research that exposed the use of intuitive, pattern matching decision making techniques of experts when facing high-stakes uncertain environments. As automation has tended to be employed to support operators when facing complex, uncertain and high-stake environments, such as flying a plane or detecting faults in a nuclear reactor, it seems sensible that with experience automated systems would be included in the intuitive and flexible decision making of operators. However, with system designers wanting to breakdown tasks to their abstract components, task complexity (Miller, 2018), flexibility (Naikar, 2018) and intuition are lost which brings into question to utility of taking such an approach. Jamieson and Skraaning (2017) argue for researchers to take a more pragmatic and inductive approach to understanding how better to support system designers in order to ensure that automated systems are built to support the operators, as the role of the human becomes more critical with more automation (Carr, 2015). From these concerns one train of thought has gained steam with researchers focusing on how to get humans and machines to work together as opposed to focusing on who performs what task (Roth and Pritchett, 2017).

As a result of this disconnect between system design and operational functions there remains reticence to move to fully automated operations, *"the aversion to algorithms making decisions that affect humans is rooted in the strong preference that many people have for the natural over the synthetic or artificial"* (Kahneman, 2011, p. 228); reticence that is not unfounded. Hoc (2000) identifies 4 failures of human-machine-interaction, barriers that are commonly cited as underpinning the disuse of automation: (i) loss of expertise, (ii) complacency, (iii) trust and self-confidence and (iv) loss of adaptability. The key focus and drive of the research into human-machine-interaction is to develop tools and systems that reduce or prevent these failures from occurring as it is these failures that underlie operators reticence to utilise new tools and system to their full extent. Nonetheless, as the nature of warfare continues to increase in complexity and the capabilities of operations are pushed to their limits, the answer to combating and remaining operationally superior is argued to lie in the appropriate use of automated systems and tools.

1.2.4 Barriers to automation uptake and use

In order to piece apart what motivates individuals to use automated systems, the barriers to or reticence towards automation use must also be explored. The literature has provided countless examples of the antecedents to not using automation. Several review papers (Parasuraman and Manzey, 2010; Parasuraman and Riley, 1997; Parasuraman and Wickens, 2008) highlight the factors that result in automation misuse, disuse and abuse focusing on the changes that can be made to the system itself in order to reduce erroneous automation usage decisions. The most commonly cited barriers are unreliability of the system, skill decrement, loss of situational awareness and complexity of the system itself. As each barrier has been thoroughly explored within the literature, only a brief overview of each will now be presented.

Unreliability of the System. Arguably the most important factor that must be considered when developing automated support tools is the reliability of the system. Research has shown that an operator's trust in the system significantly reduces when the machine errs (Dzindolet et al., 2003; Moray et al., 2000). Trust has been linked to the use of complex technological systems (Lee and See, 2004), therefore the reliability of the system arguably directly impacts upon automation usage decisions. It is not just the reliability of single aspects of the system that individuals perceive; research has shown that individuals view the system as a whole. Therefore, if one aspect of the system errs, the perception of the reliability of the system as a whole reduces. The concept of system-wide trust was proposed by Muir and Moray (1996) in their experiments in a system-supervisory context (i.e. in the context of the operator acting as supervisor to the system). The potential impact of holding a system-wide view could be disastrous. For example, if a single gauge errs but goes unnoticed with the operator continuing to hold the reliability of other gauges as a global measure of system function, a nuclear reactor could melt down, as occurred at Three Mile Island. Conversely, if the last time a GPS tool provided an updated and faster route that was in fact not faster, when it again suggests a re-programmed route, individuals are more likely to ignore this new route or even use a different GPS tool.

However, highly reliable tools and systems do not always improve performance (Dzindolet et al., 2010), and can lead to automation complacency and bias. This leads to the requirement for systems to allow the operator to remain cognitively involved in the task, particularly with regards to decision support tools (Kaplan et al., 2001). Automation complacency and bias can also result in the next two barriers to erroneous automation usage decisions, skill decrement and loss of situational awareness.

Skill decrement and Loss of Situational Awareness. When the operator‘s role becomes that of system supervisor, skill decrement (Casner et al., 2014; Wiener and Curry, 1980) and loss of situational awareness (Endsley, 1987) can occur. Two factors that are attributable to out-of-the-loop performance (Endsley and Kaber, 1999). One way in which system developers have attempted to address this barrier is to explore the appropriate LOA to task requirement. Endsley and Kaber (1999) in an experimental dynamic control task found that individuals were hindered if the automated system provided high cognitive functions. They did not however repeat the findings of Endsley and Kiris (1995) in relation to reduced SA when the operator was paired with full automation compared to a system with an intermediate LOA. The authors posit their findings may be due to the participants only performing a single task. Vigilance decrement has been observed with studies that require the participants to complete several tasks simultaneously (Parasuraman et al., 1993). It should be noted that RN personnel are involved in several tasks simultaneously during above water operations. Therefore, the potential for skill and vigilance decrement to occur, should the operators become complacent or bias towards a fully automated system, exists.

From conversations with Subject Matter Experts (SME) around the views and attitudes naval personnel hold towards semi-automated and fully automated systems, it is clear that the potential loss of skill and SA are considered as particularly salient. Arguably the fear of skill and vigilance decrement is the predominant cause behind reticence to utilise new systems to their full capabilities.

Complexity of the System. Finally, the third barrier to effective automation usage decisions is the complexity of the system itself. Poor design and lack of training can negatively impact upon the safety of the operator and the team around them (Stowers et al., 2017). Unanimously across the literature is the requirement that automated decision support tools do not operate as “black-boxes”. In order for the human operator to remain in the loop they must be able to comprehend how the automation is reaching a decision or suggested action; system designers must aim to ensure that the system operates in a transparent manner (Colebank, 2008; Johansson, 2010). Transparency also has implications for the ethics associated with using automated systems, particularly if the automated system is making decisions. One such way in which to improve the transparency of the system, and thereby in effect its simplicity, is to program the system to follow the logic that a human would use to solve the problem or reach a decision (Hawley et al., 2005). However, achieving this requires that the cognitive functions of the operator must be understood in detail which has been approached by employing cognitive task analysis methods. There may however be instances where the logic employed by the operator is flawed, for example negatively influenced by incorrect heuristics or biases, therefore the system will need to function in a cognitively different manor and yet remain understandable to the operator.

Further, research has highlighted that individuals hold a "automation schema", that results in expectations towards how the system will perform and function. Rice and McCarley (2011) explored the impact of different error types (false alarms and misses) on performance of an x-ray baggage screening task. Participants found false alarms to be more salient and therefore weigh into their decision to use the system more that automation misses. Their findings support the argument for a perfect automation schema that is held by individuals (Dzindolet et al., 2002) that assumes an automated system will be reliable. What this means is that unless informed otherwise, individuals expect automation to function in a highly reliable way and when this expectation is not met they will disuse the system. That is, participants were not expecting the automation to falsely identify objects as this contradicts the perception that automated systems function perfectly. When the system does not function as expected the perceived complexity of the system increases, reducing trust levels and thereby reducing system use.

The complexity of the system also links into the perceived use and perceived ease of system use, the key components of the Technology Acceptance Model, TAM (Davis, 1989). These components are posited to predict behavioural intention. Primarily experience with the system, allows an understanding of the perceived use and increases the ease at which the system is used, positively influencing behavioural intention and usage (Kim and Malhotra, 2005). Additionally, perceived ease of use has been found to exhibit a stronger influence than perceived use on automation acceptance (Ghazizadeh et al., 2012), further highlighting the salience of system complexity.

Summary of the barriers to appropriate automation usage decisions. These three barriers to effective and appropriate automation usage decisions further highlight the complex relationship humans have with automated systems. Although such systems can improve performance, thereby benefiting the operator, research has shown that individuals still hold a preference towards manual task completion (Navarro and Osiurak, 2015). Both system features and features of the operator must be explored and considered when building and implementing automated decision support tools (Byrne, 2018). It is clear that a broad approach is required in order to fully understand the complexities behind automation use. Additionally, it is not only the barriers to automation use that must be explored but also how incorrect decisions are made, therefore this thesis will also draw from decision making literature.

1.2.5 Overview of the literature on decision making

Cognitive processing, styles and strategies are woven into the ability to make decisions. Seminal work conducted by Tversky and Khaneman in the 1970s revealed that humans are prone to make mistakes or make less than perfect decisions (Baron, 1993). Klein and colleagues further showed the role of recognition and intuition behind decision making, particularly in challenging environments. That is, environments characterised by time pressure, uncertainty and risk (Klein, 1998). The use of biases, heuristics and time saving satisficing strategies is now commonly recognised within the literature on decision making, not only on an individual level but also within teams. Of most relevance to military tasks, and when acting under uncertainty, are the heuristic strategies of availability, representativeness and anchoring (Tversky and Khaneman, 1973; Williams, 2010).

The availability heuristic posits that individuals estimate the frequency of an event based on their ability to retrieve the recollection of such an event or, associations of the event, from memory, i.e. on the availability of that memory (Tversky and Khaneman, 1973). This is not a surprising finding as *"ones judgements are always based on what comes to mind"* (Taylor, 1982, p. 199). Research has shown that individuals are more likely to make frequency judgements based on recalled content when the task was perceived to be of high personal relevance (Rotliman and Schwarz, 1998). When facing a possible air threat that would result in potential destruction of the ship that the operators are on, it is easy to assume this task would be considered of high personal relevance. In addition to this, knowledge accessibility may reveal an innate preference for subjective recall experiences over that of accessible declarative information (Schwarz and Vaughn, 2002). This could result in error biases as incoming information may be overlooked or perceived to be irrelevant if it does not fit with the expectations associated with the recalled similar scenario.

The second heuristic discussed is representativeness, used to evaluate the probability of an object belonging to a certain class of objects, i.e. how representative is object A of the objects within class B (Tversky and Khaneman, 1973). When entering into the next stages the air defence task, that is, the identification and classification of objects, it would not be uncommon to see this bias in operation. However, research has found that individuals are prone to certain biases with this heuristic which can result in incorrect judgements being made. It is here that the development of systems, that can be programmed to statistically work out probabilities, can provide humans with support to facilitate accurate classification judgements.

Anchoring has been argued to explain why our judgements tend to be influenced by initial perception, i.e. the primary information or perception held acts as an anchor, from which all subsequent information is compared to. This can however, lead to biases with how further information is valued in terms of its salience (Baron, 1993). Research has shown that when experiencing high cognitive load or stress individuals are less able to adjust their initial disposition in light of new information compared to those who are not experiencing this cognitive load (Epley and Gilovich, 2006; Starcke and Brand, 2012). The development and maintenance of SA requires constant updating and readjustment to newly incoming information. It is obvious that cognitive load can have a detrimental impact upon the ability to effectively adjust perceptions in light of new information, and further upon the overall mission objectives.

Additionally, default bias could play a role in the continual use or disuse of an automated system. When facing complex decisions individuals have a tendency to defer to the default option, thereby not making a decision at all (Redelmeier and Shafir, 1995). Decision inertia, i.e. the failure to execute an action, is posited to occur in multi-team systems when facing a lack of clear strategic direction in a non-time bounded environment (Alison et al., 2015). Utilising NDM methods to explore decision making in a multi-agency emergency services response to a critical incident, Alison and colleagues (2015) identified three barriers to effective decision making; (i) a non-time bounded situation, (ii) multiple agencies (teams) and (iii) unclear strategic direction. Arguably all three of these barriers could exist within the air defence task. For example, prior to a critical situation, operators are working in a non-time bounded situation. Although primarily working within a single team, i.e. the team on their ship, they may also be operating as part of a task force, thereby working with multiple teams. With the increasing interoperability of allied militaries, the task force may also span several nations. Additionally, automated tools and systems can be conceived as members of the operations team. Therefore, the multi-team system on naval vessels can be comprised of human and machine teams, as well as teams from multiple nations. Finally, although the Command and Control (C2) structure is clearly defined, to avoid unclear strategic direction, with any team, when operating in critical environments there is a chance for this strategic direction to break down, or become unclear.

Operational environments present individuals with challenges to overcome in order to make effective decisions. Additionally the organisational environment can influence decision making. Drawing from an interactionist model of organisational ethical decision making proposed by Trevino in 1986, several factors from this model arguably influence automation usage decisions. Defined as *"the common set of assumptions, values and beliefs shared by organizational members"* (Trevino, 1986, p. 611) organisational culture can influence decision making of an individual in several ways. In relation to the use of automated decision support tools, organisational culture may indirectly effect automation usage decisions via obedience to authority, a normative structure, accountability and reinforcement.

For instance, new naval recruits are being trained by very experienced personnel, who were trained to carry out their tasks manually or with limited automated support, due to the type of systems that were in service 20-30 years ago. There is also the requirement for each recruit to be able to complete such tasks manually, to prevent skill decrement. However, if the focus from trainers is on the ability to manually complete such tasks this may result in maintaining the reticence to utilise automated systems to their full capability thus, forming the default option to task completion as manual. A focus on manual task completion may also extenuate the availability bias. If newly trained operators are primed to perceive total manual completion of the task as expected and desired, this may derail the decision-making process to focus on manual completion, when at times it may be more appropriate to work synergistically with the automated system. Additionally, as the operator remains legally responsible for the consequences of any decision or action taken, this further increases reticence to use automation systems. The normative structure of naval operations is maintained through traditional training and reinforcement of the history of the RN. Not that tradition and history should be lost, but arguably the organisational culture of such military organisations must adapt to future warfare, this adaptation must be at an organisational level to truly bring about change. Default bias and decision inertia have also been linked to organisational structures and climates (Eyre et al., 2008a), in particular high-accountability environments (Alison et al., 2010). Hierarchical organisations have a tendency to foster a blame culture, which increases the accountability felt by the individuals with the organisation. The military is grounded in hierarchy therefore, decision inertia may also be an influential factor that can lead to automation misuse or disuse.

Dzindolet et al. (2010) posit that automation misuse and disuse may result, in part, from automation bias. Defined by Mosier and Skitka (1996) as *"the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing"* (p205). Automation bias can result in appraisal errors - which are errors in judgement in relation to how best to complete the task at hand, i.e. with an automated or non-automated option; and/or intent errors knowing the utilities of the automated system and yet not taking into account these properties when deciding on using the system or not (Beck et al., 2007). Both appraisal and intent errors can negatively impact upon automation usage decisions, i.e. an individual holding an automation bias may misappraise the situation (an appraisal error) and use the automated option when the non-automated option was more appropriate for task completion (an intent error). Several aviation disasters in recent years have been due to the crews' overreliance on their aircrafts automatic pilot and have ultimately resulted in the loss of life of the crew and passengers. For example, in July 2013 Asiana Airlines flight 214 crashed attempting to land at San Francisco, of the 291 passengers, 3 were killed and 187 injured, 49 seriously (Konnikova, 2014). The cause of the crash was ruled due to the pilot selecting the wrong autopilot setting and then failing to recognise that the aircraft was going too slow and descending too fast. Conversely, not utilising an automated tool in favour of manual control can also have negative consequences. For example, the Costa Concodria disaster in January 2012 resulted in the loss of life of 32 passengers. The captain chose to manually control the navigation of the vessel, which upon investigation following the crash was found to have been taken off the computer programmed course, resulting in the cruise ship hitting a coral reef (Levs, 2012). Therefore, the ability to discriminate when to rely on an automated system and when to rely on manual control is critical to successful task completion.

It has been posited that involving individuals in the development of an automated aid can enhance reliance (Kaplan et al., 2001). This study also highlighted the value of allowing decision makers to remain cognitive involved in the decision process. Supported by Casner et al. (2014) who found that 'think aloud' protocols mitigated pilots cognitively disengaging from the task when using an automated flight path system. These studies show how easy it is for individuals to disengage from a task when they assume that the automated feature is performing correctly. However, the 7th May 2016 saw the first fatality from an automated car[5]. The autopilot was unable to discriminate between the white cloudy sky and a white trailer truck and did not take avoidant action. It should be noted however, that although in autopilot, the driver was still required to remain alert in order to take back control if necessary, as is expected within human-automation teams. Unfortunately, the driver in this instance did not take back control in time to avoid the crash, and ultimately lost his life. A second incident that involved an autonomous car killing a pedestrian on 18th March 2018 has also raised questions on the safety of such vehicles[6]. Later analysis of the incident showed that the car incorrectly categorised the pedestrian as a bicycle and was not programmed to perform an emergency brake procedure. The system designers were relying on the individual in the car, who was not paying attention to the road at the time of the collision, to stop the vehicle [7]. Incidents in the aviation industry, as well as those just

[5]https://www.teslamotors.com/blog/tragic-loss

[6]https://www.theguardian.com/technology/2018/mar/22/video-released-of-uber-self-driving-crash-that-killed-woman-in-arizona

[7]https://www.wired.com/story/uber-self-driving-crash-arizona-ntsb-report/

mentioned, also highlight the importance of not assuming that the automated systems in use can be easily understood by the people operating them. For example the Lion Air Boeing 737 MAX 8 incident in October 2018. The pilots were overwhelmed by the feedback they were receiving from the system during an error which resulted in them not following a checklist procedure to recover the aircraft. Ultimately 189 people were killed when the plane nosedived into the sea (Learmount, 2019).

Arguably, the drivers, pilots and captains in these cases succumbed to automation complacency, as has been seen with many incidents across the aviation industry and within the military. Operators of automated systems are required to continually re-evaluate and update their understanding of the environment, as they would normally do when completing a task. However, an additional evaluation of the automated system must now also occur alongside evaluating the environment in order to supervise the functioning of the system. If overreliance on automation leads to complacency then individuals will not understand the current situation and therefore will not be anticipating potential future actions, decisions or consequences. Therefore, research is required to explore what makes some individuals more susceptible to automation complacency than others and how best to mitigate the negative consequences of such complacency.

1.2.6 Naturalistic decision making and decision centred design

NDM research has sought to explore cognitive work performed within complex sociotechnical environments and as a research domain has grown from work conducted with the US Military and Navy (Schraagen et al., 2008). NDM research has been defined as the study of how experience is used by people in the field to support their decision making (Zsambok and Klein, 1997). This body of literature has contributed greatly to furthering our understanding into, but not limited to, how complex decisions are effectively made within military and emergency services contexts. NDM approaches are pragmatic (Gore et al., 2006) and have been well established within the literature, dating back to the late 1980s when Gary Klein and colleagues (1986) explored the decision making of fire ground commanders. The development of the Recognition Primed Decision making (RPD) model, that depicts the decision-making process that occurs in extremis, opened the field of decision making to look beyond rational decision-making theories i.e. utility theory. The RPD model depicts how experience facilitates the ability to intuitively respond to a situation through utilising subtle cues within the environment (Klein, 1998), highlighting experts tacit knowledge of the task. Experts use mental simulation to evaluate the worth of a particular course of action, often only considering one course of action at a time (Ross et al., 2006).

Johnston et al. (1997) found U.S. Navy personnel utilising more hypervigilant decision-making strategies (i.e. selectively scanning information available to rapidly attend to the meaningful data) performed significantly quicker, without a cost to their accuracy, whilst also making a significantly greater number of accurate target identifications compared to those using vigilant decision-making strategies (i.e. those who conducted a systematic information search and considered multiple alternatives). Further, Ross et al. (2004) found that military personnel did not find training in the RPD model to be anything new, that is, they continued to formulate and make decisions as they naturally did. Conceptualised by Boyd in 1987 OODA loops (Observe, Orientate, Decide and Act) remain the dominant model behind C2 development and military decision making (Ministry of Defence, 2010). Although not intended as a model of decision making (Brehmer, 2005), OODA loops provide a useful tool to explain the high-level stages of thinking that individuals cognitively progress through to achieve their goal. Each stage of the loop can be iterative, with individuals completing multiple OODA loops sequentially and in parallel. A key component of being able to progress through the loop is to understand the current situation, this enables you to anticipate the next action or decision. OODA loops are similar in nature to the RPD model citing intuition and experience as key facilitators of effective decision making when faced with time pressured and critical environments.

The value of understanding the macrocognition of experts has also been argued to facilitate the development of better technology interfaces and training programmes (Klein et al., 2003), exploring where technology can support macrocognition of individuals and teams and where it may hinder it. This body of literature has underpinned the development of decision centred design (Hutton et al., 2003) which uses NDM methods such as cognitive task analysis to build an understanding of the experts task and where a system may be of use in supporting that task, for example, the Tactical Decision Making Under Stress (TADMUS) programme (Cannon-Bowers and Salas, 1998). TADMUS had several aims, one of which was to develop a decision support system that would enhance tactical decision making in a single ship conducting the air defence task in a littoral environment. Taking an NDM approach to this programme the researchers aimed to create a decision support tool that "cognitively fit" to the way in which operators completed the task. Upon testing the interface experts were found to identify more critical contacts earlier when using the decision support system suggesting that automated system which support the intuitive process of feature matching as well as the more analytical process of explanation based reasoning can provide effective support to operators (Morrison et al., 1996).

A key premise of NDM research is to talk to, and crucially, to learn from experts. This is due to expertise being associated with superior decision making in certain environments as it provides individuals will high-levels of proficiency and task based knowledge (Kobus et al., 2001). This proficiency facilitates the speed at which good judgements can be made or actions completed (Hoffman et al., 2013). However, when exploring sociotechnical systems, Mosier et al. (1998) found that pilots with more experience made more automation errors than pilots with less experience. This suggests that it is not just expertise in the task that is required by personnel operating in sociotechnical environments but also expertise of the system(s) they are interacting with. Therefore, focus remains on how to facilitate the development and maintenance of expertise (Gore et al., 2006). It has been found that both novices and experts benefit from slow and deliberate thinking (Moxley et al., 2012). Accordingly, slow and deliberate practice with the system(s) they are required to co-operate with may provide human operators the chance to develop expertise on how the system functions.

Taking an NDM perspective provides researchers with a robust way in which to approach exploring the cognitive processes associated with automation usage decisions. However, as the very nature of automated systems stems from the development of complex algorithms, this thesis will also draw from the heuristics and biases literature. According to the NDM school of thought, expertise and intuitive judgments arise from experience, whereas, from a heuristics and biases perspective, intuitive judgements arise from the use of simplifying heuristics (Kahneman and Klein, 2009). Therefore, when making decisions the human-machine dyad (of operator and automated system) conceptually utilise both experience and heuristics/statistics to make judgements. To support the development of this human-machine dyad the human operator must learn the task skill and the skills required to use the system effectively in order to prevent incorrect intuitions. In line with this there is a danger in being overconfident in ones own abilities which could lead to missing a novel cue that has not been seen before (Klein et al., 2011). It is in these instances that working alongside an automated system may mitigate against the error of missing a critical cue as the system may flag the error. However, the system is only as good as the information it has been coded and trained on and so the problems that can arise with novel events or cues may continue to persist. The introduction of automated systems has in parts reduced the challenges associated with making decisions in military environments, however, in other ways automated systems have also increased the complexity of these decisions (Brown, 2007).

A key factor revealed by NDM research that can derail decision making is uncertainty. Lipshitz and Strauss (1997) classified uncertainty into two categories, endogenous - relating to the problem environment and exogenous - relating to the system responding to the environment. In relation to human-machine-interaction, endogenous uncertainty can be conceived of as the incompleteness of the incoming data and the uncertainty around how reliable the source of that information is. Exogenous uncertainty can be thought of as the relationship between operator and system, for instance, how much trust the operator has in the system to inform them of the relevant information and how reliable is the information being presented to the operator from the system. Both forms of uncertainty have been found to negatively impact upon the decision making process. Endogenous uncertainty was found to have a greater impact during the SA and plan formation phases of a decision whereas exogenous uncertainty was found to have a higher impact upon the plan formation and action phases (Alison et al., 2014). Uncertainty can result in the inability to make a decision (Shafir, 1994), individuals can find themselves trapped in a loop of continually searching for more information, which can detrimentally impact upon performance (Klein, 2015). Inevitably, within the air defence task, this would result in an object not being classified which could be potentially fatal if that object was in-fact a hostile unit about to strike.

> "each Commander can only fully know his own position, that of his opponent can only be known to him through reports, which are uncertain; he may, therefore, form a wrong judgement with respect to it upon data of this description, and, in consequence of that error, he may suppose that the power of taking the initiative rests with his adversary when it lies really with himself"[8]

The increasing capabilities of sensors have to some extent reduced information uncertainty. It is now possible to detect, record and log far more data points than historically was the case. Automation applied to the fusion of these types of data enables the operators to engage in higher level tasks and decision-making processes, such as threat evaluation. However, as Clausewitz wrote and is commonly accepted, military environments are inherently uncertain (Kobus et al., 2001). It is how this uncertainty is dealt with that can make the difference between a successful mission and a failure.

Although technology can reduce certain aspects of uncertainty, simultaneously the inappropriate application of systems can increase uncertainty relating to the system itself (e.g. the system operating as a "black-box"). It is vital to ensure the appropriate level of automation is used for the task in question. For instance, Kaber et al. (2005) found that using adaptive automation during the information and acquisition stages of the decision-making process improved performance and reduced workload. However, automation applied to the information analysis and decision-making stages (i.e.

[8] Von Clausewitz (1832) (p.115).

the high-level stages of the decision-making process) resulted in increased workload. Although the prescription of levels to a task facilitates comprehension of where the system will slot into the decision-making process, artificially deconstructing tasks into subcomponents has led to the loss of flexibility which is required when facing uncertain environments (Naikar, 2018). Therefore, research should explore approaching the development of automated systems in a new way, one that embraces the complex and at times messy way in which teams perform tasks.

More recently there has been a call for a more pragmatic approach to researching human-machine-interaction and not just subscribing to narrow models of design or conceptualisation (Jamieson and Skraaning, 2017). As Stowers et al. (2017) write, *"it is essential for attitudes, behaviours and cognitions to be continuously monitored and quantified, as they provide great insight into how to improve or design more usable, safe and efficient systems"* (p177). Additionally, the development and application of automated systems must move beyond looking primarily at workload as a user assessment measure (Schwarz et al., 2014). Therefore, this thesis proposes to take a pragmatic and holistic approach to exploring automation use in the operational setting of above water warfare. Further, it is not just cognitions towards automation and/or the specific automated decision support system that can influence uptake and task performance. Cognitive traits specific to the end user may also play a role in automation usage decisions. Hoff and Bashir (2015) highlight the requirement for further research into the role of cognitive factors behind the development of dispositional trust, and therefore, automation usage decision. This thesis proposes to fill this gap, exploring the influence of six cognitive traits and their association to automation usage decisions: (i) Cognitive Flexibility; (ii) Need for Closure; (iii) the self-regulation processes of Assessment & Locomotion; (iv) Impulsivity; (v) Conscientiousness; and (vi) Propensity to Trust.

1.2.7 Thesis Structure

This thesis aims to add to the body of literature exploring how to improve maritime capability. The focus of this work is on exploring the operational application of automated systems in the maritime above water environment and begins by describing the environment naval personnel are operating in. Next this thesis will explore current RN use of automation and where personnel view that use to extend in the future, evidencing if current personnel are consulted in the development of new automated systems that they will be required to use. Finally, this thesis will present the development of an above water micro-world used to explore individuals uptake of an automated decision support tool in a holistic way.

This chapter has discussed the psychological underpinnings of this thesis, providing context on the problem space and an overview of the historical literature that has been conducted on human machine interaction. It has drawn from the NDM and heuristics & biases literature on decision making in critical environment as the use of automated systems in military contexts impacts directly upon decision making. Additionally, the factors that effect decision making in these environments arguably also impact upon decisions that relate to 'if to' and 'how to' use automated systems.

Chapter 2 presents the mixed-methods approach taken by this research, discussing the challenges that are characteristic of mixed-methods and interdisciplinary research. The merits of applying NDM approaches in complex, high-stake environments that are often interdisciplinary are also discussed. This thesis consists of three stages which are presented in this chapter. Stages 1 and 2 employ qualitative methods to collect rich data from subject matter experts within the RN. Stage 3 concerns the development of the Automatic Radar Classification Simulation (ARCS) which was designed by the author of this thesis to provide academia with a high-fidelity microworld used to explore automation usage decisions when performing a threat detection task.

Chapter 3 sought to qualitatively explore the decision stages to the air defence task. To begin with is the introduction of the historical research on the air defence task, specifically identifying two key studies on this task by Holt in 1988 and by the TADMUS programme which began following the USS Vincennes incident. Vignette questionnaires were completed by (N=7) RN personnel with the data being analysed thematically to reveal three high-level stages to the air defence task: Observe; Identify and Classify; and Decide and Act. Occurring iteratively, these stages can be facilitated by effective communication and situational awareness but also hindered by cognitive pressures, ambient situational factors such as uncertainty and by organisational constraints for example, ROE. The key finding is that the stages of the air defence task have not significantly changed in the last 30 years. The areas identified that could benefit from the introduction of automation systems remain the same. This brings into question if systems have been employed and are used ineffectively by personnel, or if systems that have been developed do not provide the support personnel require.

Chapter 4 presents the second stage of this thesis which sort to explore the current state of automated systems and how they are used operationally by RN personnel. Potential future use of automated systems as well as opinions of current RN personnel towards the existing procurement process were also explored. A qualitative questionnaire was completed by (N=46) currently serving RN personnel and as with chapter 3, thematic analysis was conducted upon the data collected. It was highlighted that automation use in ubiquitous in daily operational environments, supporting the current and future doctrine that cites the importance of automated systems. However, a critical concern was uncovered with the lack of engagement between personnel and system designers, which results in systems not being built fit for purpose. The implications of these findings are discussed alongside recommendations for how to bridge the gap between personnel and system designers.

Chapter 5 describes how ARCS was initially tested with two pilot studies (N=6 and N=5 respectively). This chapter will present the adaptable features that have been built into the ARCS microworld such as, the ability to edit the functionality of the system, adapt the scenarios to test different research aims and the ability to incorporate questionnaire metrics into the task. The results of the pilot test confirm the ability of ARCS to collect a vast amount of data on an individuals performance, their use of the system and the rationales behind their decisions. This chapter will present an overview of the data that can be collected using ARCS and how a holistic approach to understanding human machine interaction is achieved by using this microworld.

Chapter 6 presents the findings from the student experiment conducted with ARCS. Students from the University of Liverpool (N=42) took part in a two-stage experiment to explore the motivational and cognitive factors that may influence the uptake of an automated decision support system. This chapter explores the findings from this study in relation to: i) the influence that prior information on the system and experience with the system has upon uptake; ii) how workload and accountability influence task performance and the decision to use the automated system; and iii) how different cognitive processing styles may modulate automation uptake and use. Quantitative and qualitative analysis was conducted on the data collected. The main finding of this chapter is that automation use does not always result in a significant benefit to task performance and that capturing a complete picture of the factors that influence automation uptakes requires research to take a mixed-methods approach.

Chapter 7 concludes this thesis by providing a summary of the findings presented in each chapter. The key findings of this thesis are that a disconnect between end user and system designer results in problems with the development and introduction of new automated systems. For example, systems are not fit for purpose and/or do not perform as expected which has negative repercussions on capability and financial pressures. Therefore, a central recommendation of this thesis is for focus to be placed on exploring ways to develop and maintain effective collaboration between and across military organisations, academia and industry. A second finding from this thesis is the value of utilising immersive environments to train and test new systems. However, there is a paucity of such environments. Thus, a significant contribution of this thesis is the development of ARCS, a novel microworld that simulates the initial stages of the air defence task. This chapter also presents how some of the challenges associated with naturalistic research were encountered and addressed, particularly in relation to ethical consideration and access to specialist expert populations. The methodological strengths and weaknesses are also discussed alongside interesting directions for future research.

Chapter 2

A methodological approach to interdisciplinary research in real-world and experimental settings

2.1 Introduction

To achieve the aims of this thesis, the adoption of a mixed-method approach is proposed. This approach provides the research with enough flexibility to include the strengths of both qualitative and experimental approaches to research.

Drawing from the methodological literature and from the researcher's experiences of undertaking this thesis, this chapter will present some of the challenges characteristic to mixed-method and interdisciplinary scholarship. This chapter will be divided into three sections that will:

(i) Discuss the challenges characteristic to mixed-methods research and interdisciplinary scholarship

   (a) Describe the framework of mixed-method design developed to provide transparency of this thesis, as well as a framework that can be adopted and further developed by future research.

(ii) Discuss how the paradigm of Naturalistic Decision Making provides a robust outlook to take towards interdisciplinary research, for example by providing methodologies that facilitate collecting data from expert populations

   (b) The Critical Decision Method will be discussed, and how key features of this method were adapted into a vignette survey to accommodate to one of the challenges to 'real-world' research: access to expert samples.

(iii) Discuss Experimental Design and Development

   (c) How the Automatic Radar and Classification System (ARCS) was designed and developed to provide academia with a high-fidelity test bed for novel ways to explore human automation interaction.

2.2 Challenges characteristic to mixed-methods research and interdisciplinary scholarship

2.2.1 Methodologies

The researcher's methodology prioritises and guides the development of the research questions (Hesse-Biber, 2010), and may implicitly lead to the selection of certain methods. Therefore, it is prudent to be transparent about the researcher's epistemology.

Primarily this thesis takes an interpretivist perspective, that is the view that multiple subjective realities exist through individuals construction and perception of their own lived reality. The nature of this thesis and the area under study is interdisciplinary. Therefore, one challenging aspect of this work is the requirement to incorporate multiple methodologies to address the research questions. With this is mind, a mixed-methods approach is deemed highly suitable for this work. A requirement of taking a mixed methods approach involves the researcher to be open to alternative views and epistemology. As Greene et al. (2001) wrote *"openness to other views and perspectives, not just to rival explanatory hypotheses but more profoundly to rival ways of thinking and valuing"* (p32). Thus, positivistic and postmodern approaches are also incorporated into this work. Hesse-Biber in her textbook on mixed-methods research wrote *the method is but the tool; the methodology determines the way in which the tool is utilized* (Hesse-Biber, 2010, p. 17). This thesis uses a combination of methods to address the research questions under exploration. Each method will be discussed in full within this chapter. However, to begin with is a discussion on taking a mixed-methods approach to this work and the reasons why this approach was taken.

2.2.2 Taking a mixed-methods approach

Within the literature several different terms have been used to denote the use of both quantitative and qualitative methods and forms of analysis. For this thesis, the following definition was applied:

"A mixed methods study involves the collection or analysis of both quantitative and/or qualitative data in a single study in which the data are collected concurrently or sequentially, are given a priority, and involve the integration of the data at one or more stages in the process of research" (Creswell et al., 2003, p. 165).

Few guidelines exist on when, why and how to use mixed-methods (Bryman, 2006) however commonly cited in the literature is Greene et al. (1989) five reasons researchers adopt a mixed-methods approach to research: (1) Triangulation, (2) Complementarity, (3) Development, (4) Initiation and (5) Expansion.

Triangulation, commonly viewed as the process of conducting *"quantitative and qualitative research [that] might be combined to triangulate findings in order that they may be mutually corroborated"* (Bryman, 2006, p. 105). Triangulation is employed within the research design to look for convergence of data collected, enhancing the credibility of research findings. Bryman (2006) reported that of the 80 articles that were coded to have used a triangulation approach, only 19 articles had articulated triangulation in their rational for employing mixed methods. However, 29 articles included in the content analysis had stated their use of multi-methods for triangulation, with only 19 of those articles being coded to have used multiple methods for the purpose of corroborating their findings. The proposed research framework in this chapter identifies triangulation as an underpinning factor in adopting a mixed-methods approach. This is both for the increased validity using similar methods can produce as well as the increased depth of understanding gained from triangulating qualitative methods (Denzin and Lincoln, 1994).

The second reason proposed by Greene et al. (1989) is that of *complementarity*. Which, according to their 1989 article *"seeks elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method"* (p.259). With this purpose in mind the adoption of mixed methods can allow the researcher a holistic understanding of the research problem. It should be noted that depending on the research problem under exploration, it may be that a complete holistic understanding is not possible through a single mixed-methods study. It may be that from the primary study springs further inquiry into the phenomena in question.

Development enables the results from one method to inform the other. This is argued to increase the validity of constructs identified through the research (Greene et al., 1989). For example, this thesis begins with a qualitative study to explore the decision-making stages Royal Navy (RN) operators go through to complete the air defence task. The results of part 1 of this research have directly informed the development of the microworld experiment (quantitative method) and survey (that encompasses both qualitative and quantitative measures).

Methods can inform each other, therefore, value from a mixed-method approach can be gained via exploring the contractions or paradoxes that may emerge between methods. This is the fourth reason for adopting a mixed-methods approach, *initiation*. The expected findings from the quantitative method may not fit with the findings from a qualitative method. This will allow the researcher to take a new perspective towards the findings, to explore the possible reasons behind the paradox or contradiction. This fresh perspective will broaden the inquiry and the depth of analysis by the application of multiple viewpoints onto the findings.

Finally, *expansion*, the use of multiple methods to *"extend the breadth and range of inquiry by using different methods for different inquiry components"* (Greene et al., 1989, p. 259). This reason is commonly utilised when evaluating programmes. Researchers will adopt quantitative methods to assess program process via standardised tests of validity and reliability. This will then be complemented by a qualitative method to explore programme outcome through in-depth understanding of the results and/or the real-world implications of the programme.

The field of human-machine-interaction is currently an important research area due to the increased sociotechnical relationships that are developing across a wide range of domains including the military, healthcare and emergency services. As highlighted in the previous chapter, in the majority of the current research into human interaction with automated decision support systems there is an emphasis from the view of the system developer. For example, how can research better articulate to the system developer's features of the system to improve human machine interaction. Yet, there remains a gap in this literature with looking more specifically at the humans using the systems. Key exceptions do exist, i.e. (Beck et al., 2007; Skitka et al., 2000; Szalma and Taylor, 2011) and the findings that have been presented produce a complex picture of human-machine-interaction. Greene et al. (2001) argue that certain types of *"complex, multiply-determined, dynamic social phenomena- can be better addressed through the multiple perspectives of diverse methods than through the limited lens of just one"* (p27). The field of human-machine-interaction is a highly complex, multiply-determined and consistently dynamic environment within which operators are having to make increasingly complex, time pressured decisions. Therefore, the adoption of mixed-methods is argued to be highly suitable for this work. Mixed-methods work goes hand in hand with the potential for unexpected findings (Bryman, 2006) and concerns have also been raised with unintended consequences of adopting mixed-methods when a single method would suffice. Similarly, the impact of the researchers or the research teams skill in the use of certain methods (Hesse-Biber, 2010) could result in findings that do not contribute to theoretical understanding. In order to mitigate against this, the framework developed for this thesis is presented.

The majority of research that has looked specifically at the philosophical stance and value of multi-methods and/or mixed-methods research has stemmed from the social inquiry literature. The current nature of psychological research presents as a domain that is encouraging researchers to look beyond their boundary of a single research area. For instance, naturalistic decision-making research which began following the work of Gary Klein and colleagues in the 1980s is an ever-growing research community that values a multi-method way of thinking (Gore and Conway, 2016; Lipshitz et al., 2001). It is this approach to conducting research, that has allowed inquiry into complex multi-faceted interpersonal and sociotechnical decision-making environments such as the military (Kaempf et al., 1993; Militello et al., 2015; Pascual and Henderson, 1997), medical (Patterson et al., 2009; Power, 2017), emergency services (Alison et al., 2015), policy makers, economics (McAndrew and Gore, 2013) and sociotechnical teams (Hutton et al., 2003). The Naturalistic Decision Making (NDM) philosophy promotes positivist research that aims to provide practitioners with practical recommendations to support their operations (Gore and Conway, 2016; Klein, 2015). Additionally, in their report on the 6,679 case studies within the 2014 Research Excellence Framework, Kings College London highlighted that 87% of these cases were multidisciplinary and led to 3,709 identified unique pathways for research to have an impact upon society (King's College London, 2015). This thesis draws on NDM philosophy in the approach taken to exploring the research questions proposed.

2.2.3 Advantages of taking a mixed-methods approach to interdisciplinary scholarship

The value of conducting mixed methods research is widely recognised (Greene et al., 2001; Hesse-Biber, 2010). Utilising multiple methods can strengthen the study by cancelling out or neutralising the limitations a single method has was used alone (Jick, 1979). To ensure quality of data collection and analysis when adopting a mixed methods approach the researcher must explicitly be aware of and discuss the paradigms each method stems from. Greene et al. (2001) write *"good mixed-method practice is achieved by thoughtful mixed-method planning"* (p29). The primary aim of this chapter is to present the planning and framework that underlies this thesis, and as such, the creation of this chapter was iterative. At each stage of the research this chapter was revisited and revised in order to ensure that the methodological framework that grounds this work is made clear to the reader. The iterative creation of this chapter mirrors the iterative nature of mixed-methods research.

Through the comprehensive literature review, presented in chapter 1, it is apparent that the area of human-machine-interaction is a complex, multi-faceted domain. It is these types of research areas that benefit from the adoption of multiple methods to bring depth and breadth to the research. Finally, to reduce the potential impact of lack of skill with certain methods or forms of analysis, the framework has been developed with methods the researcher has experience of conducting. Where a novel form of analysis or aspect of method design, for example, with the coding of the microworld a computer programmer was consulted and collaborated with the building of the microworld used in stage 3 of this work.

2.2.4 Framework of the mixed-method approach developed for this thesis

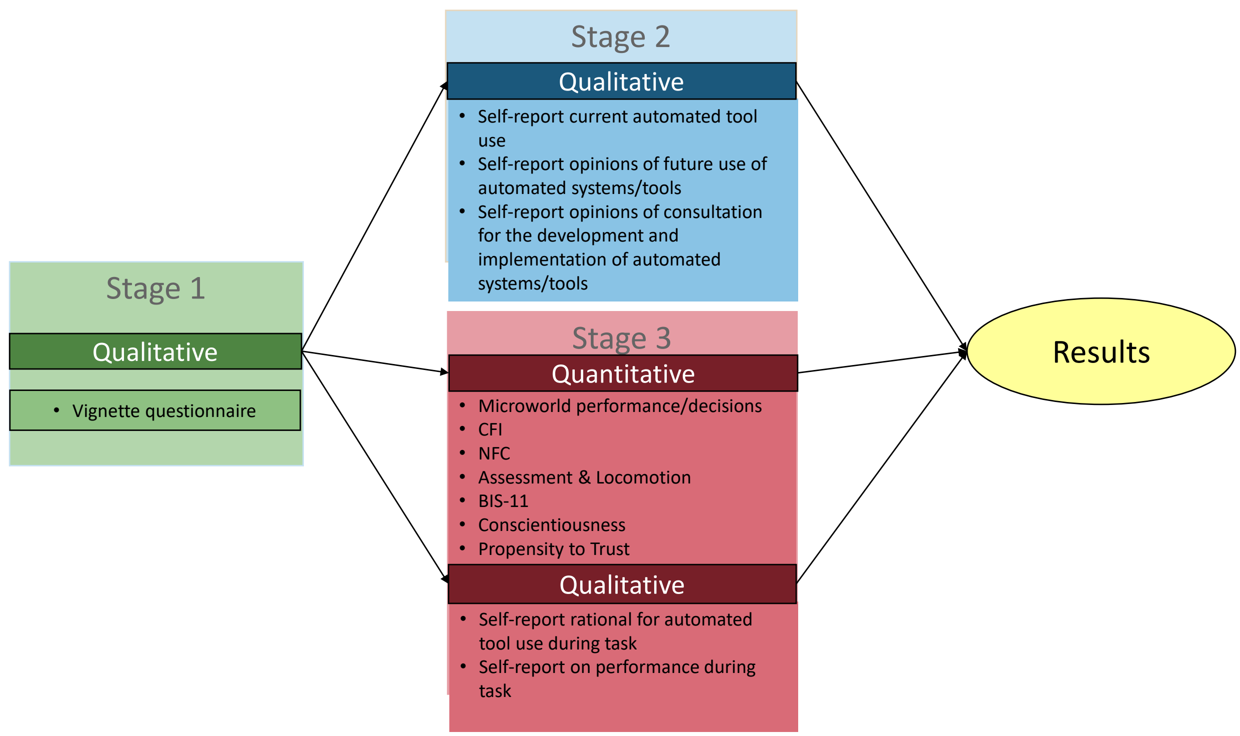


Figure 2.1: Visual Representation of Procedures

Figure 2.1 depicts the visual representation of procedures and the proposed measures used at each stage within this framework. A further consideration for planning mixed-methods research is when integration between the methods will occur. The framework developed for this research planned data integration at the problem/question stage, data collection and interpretation stages (see Figure 2.2). This framework has drawn from the commonly practiced use of a short qualitative measure as a pilot study (Creswell et al., 2003). Stage 1 of this framework was designed to enable the researcher to gain insight into the research problem and domain. A qualitative based survey was developed to generate rich and in-depth conclusions from a small sample of Subject

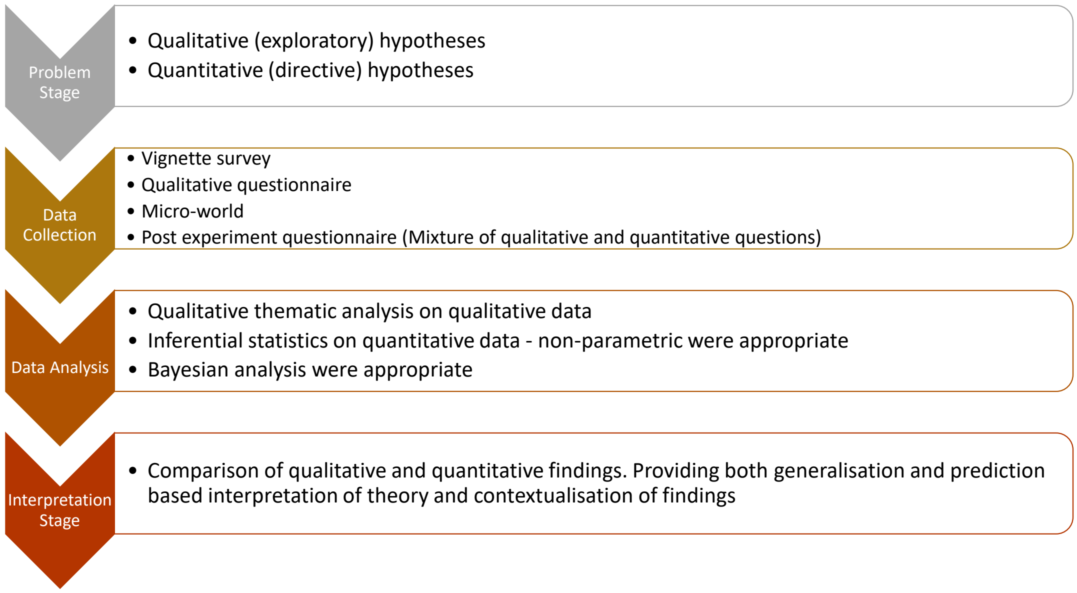


Figure 2.2: Integration of methods at each stage of experiment 2 and 3

Matter Experts (SME) within the Royal Navy. The use of small, select samples is a key feature of real-world research. Researchers face a challenge with balancing the ability to access these small select sample groups (e.g. within the emergency services, governments and military) and the 'power' of their studies. Stage 1, combined with extensive conversations with experts, enabled the researcher to gain quasi-expert status through immersion in the practitioners world (Pfadenhauer, 2009). This knowledge directly fed into the development of stages 2 and 3.

Stages 2 and 3 were designed to answer the research hypotheses that stemmed from stage 1, a comprehensive literature synthesis and from the researchers knowledge. To improve the 'power' of the work the framework is divided at this stage to include a select sample of SME (stage 2) and a student sample (stage 3). It is argued that by taking this dual approach to data collection it is possible to maintain the representativeness and the power of research findings to a generic population. Whilst, through stage 2, providing a direct link between generic and specific populations. This approach also mitigates the recurring issue with real-world research, one of access to experts and practitioners.

2.2.5 Conclusions of the challenges typical to mixed-method and interdisciplinary research

This chapter has, so far, outlined some of the typical critiques and challenges associated with mixed-method and interdisciplinary research. The framework presented provides transparency for how, and more critically, why certain methods were chosen at the onset of this work and at the data collection, data analysis and interpretation stages. This framework addresses the common challenges associated with mixed-methods work; that of the researcher identifying where methods integrate and how to achieve this robustly. To continue this transparency, the next section of this chapter will specifically focus on the methods adopted within this thesis.

2.3 Interdisciplinary Scholarships: why NDM provides a robust approach to research

NDM methods have grown from research with the US Navy and Military (Schraagen et al., 2008). Providing a methodological paradigm from which to conduct research into complex sociotechnical environments and systems. This paradigm provides researchers with a good grounding to conduct interdisciplinary research due to the development of shared language and methods (Robertson et al., 2003). The development of this paradigm suits the nature of interdisciplinary scholarship, providing adaptive and flexible methods to explore complex realities. Research into human-machine-interaction has taken many forms, from student samples (e.g. Osiurak et al., 2013; Rice and McCarley, 2011) to research with experts (e.g. De Greef et al., 2010; Kaplan et al., 2001; St. John et al., 2005). Yet, there remains a call to explore new and different methods of testing decision support systems (Todd and Benbasat, 1994) as challenges with how tools are transferred from sterile research environments into the real-world without falling into the camp of brittle technology remain. A key tenet of NDM research is to explore choice implementation in the real-world (McAndrew and Gore, 2013), therefore, members of ergonomics and human factors communities have already begun to identify the strengths of taking an NDM approach to human-machine-interaction, for example adopting a decision centred design approach. NDM methodology promotes the crucial role of situational awareness, dynamicity, uncertainty, and mental imagery in how decisions are made in operational environments (Lipshitz et al., 2001). NDM takes a positive approach, focusing on what practitioners and experts do well, *"seek[ing] to reduce mistakes and thereby seek to help decision makers perform skilfully and use their experience and intuition effectively"* (Gore and Conway, 2016, p. 332). The context of this thesis focuses on above water maritime operations, a decision-making environment conducted by experts in Naval services across the world. Taking an NDM approach allows this thesis to explore and provide valuable recommendations for practitioners operating in these environments.

A collection of NDM methods that have been used successfully in previous research to explore decision making are cognitive task analysis methods. The primary aim of cognitive task analysis methods are to yield information about goal structures and cognitive processes that underlie observable task performance (Chipman et al., 2000).

Flanagan (1954) argued that the critical incident technique provides researchers with *"a flexible set of principles which must be modified and adapted to meet the specific situation at hand"* (p336). Cognitive task analysis has been applied to a wide range of domains and specific decision-making situations or tasks, evidenced by the development of the Applied Cognitive Task Analysis method (ACTA). ACTA was developed to translate cognitive task analysis methods from the research community to operational contexts (Militello and Hutton, 1998). The premise of NDM research is that to improve operational environments or functioning the first step is to understand how experts perform the task under observation (Klein et al., 1989). The variety of organisational domains that can be explored utilising NDM methods and the individual challenges associated with each domain, and within that, each specific decision-making point/sequence/task require researchers to adapt methods to suit the task at hand. This thesis focuses on above water maritime operations, tasks that are carried out by a limited number of experts within the UK. Additionally, access to these experts is hindered due to their full-time requirement to the job that takes them all over the world. Therefore, in order to facilitate the collection of data from this novel and restricted group of experts, questionnaires were adapted from cognitive task analysis approaches.

2.3.1 Stage 1 - Adaptation of the critical decision method into a vignette scenario

The Critical Decision Method (CDM) interview aims to *"identify the knowledge requirements, expertise and goal structures involved in performing a decision makers work"* (Wong, 2003, p. 327) and to generate relevant recommendations for supporting high stakes decision making (Crandall et al., 2006). This interview method has been effectively used by research exploring high stake decision making that spans domains such as; emergency services (Klein et al., 2010; Power, 2015), medical staff (Crandall et al., 2006) and the military (Shortland and Alison, 2015). This method enables practitioners to "tell stories", facilitating an incident-centric approach to identifying critical decision-points. Drawing knowledge from expert samples is possible due to experts holding the meta-cognitive abilities to describe their knowledge as well as to posses it (Klein and Militello, 2005). Therefore, although there are additional challenges to the researcher in terms of accessing experts, the development of interpersonal relationships and the balancing of research aims with the practicalities of conducting research, access to even small samples of experts can provide a wealth of highly rich and contextualised data and knowledge.

The CDM interview begins with practitioners deciding on an incident they would like to recall, one that would be suitable for the specific decision environment the researcher is exploring. For example, if the research focus is on when decision inertia may occur the participant would be asked to recall a situation in which they faced a very difficult choice between two potentially bad outcomes. Following the initial recall of this occasion, the researcher moves into the timeline and deepening stages of the interview. These stages guide the practitioner to work through each stage of the decision process, what information was available, what actions they took etc. Essentially, these stages flesh out the key decision features and points. The researcher is then able to utilise aid memoir prompts (examples in Table 2.1) to discuss the decision from different perspectives, identifying goals and priorities and the experts overall opinions of the decision-making process. For a full breakdown of the stages to the CDM interview see (Crandall et al., 2006).

Table 2.1: Examples of aide memoir prompts that can be used within the CDM adapted from Klein et al. (1989) and Shortland (2017)

Cues:	What were you hearing/thinking/noticing during this situation?
Information:	What information did you use in making a decision or judgment?
	What did you do with this information?
Goals and Priorities:	What were your specific goals and objectives at this time?
	What was the more important thing for you to accomplish at this point?
Assessment:	If you were asked to describe the situation to someone else at that point, how would you describe it?

Due to the restrictions to accessing SME within the RN, a 4-section vignette scenario questionnaire was developed to elicit similar knowledge to that extrapolated from conducting CDM interviews. The limitations with conducting questionnaire based research is acknowledged, particularly when compared to an in-depth knowledge elicitation interview technique. However, due to the area under research within this thesis, the limited access to SME time and the very nature of their job roles and locations, the development of an online vignette questionnaire was deemed to be the next best option to elicit this knowledge. The decision to develop vignette scenarios as the basis of the questionnaire also aimed to improve the depth and quality of the qualitative data gained from SME. Vignettes enable the exploration of normative issues in a way in which reality, and the complexities associated with it, can be approximated (Finch, 1987).

Defined by Evans et al. (2015) as *"a brief, carefully written description of a person or situation designed to simulate key features of a real world scenario"* (p.162), vignette scenarios have been successfully used to explore and investigate decision making with practitioners in settings that are difficult to assess experimentally. It is noted that, a long-standing concern of the use of vignettes is how realistic the scenarios presented are. To mitigate this, and in order to develop realistic scenarios that are internally valid, scenarios were developed in collaboration with a member of Dstl staff with a Military Advisor (MA) role. Therefore, the terminology used for each scenario was designed to be understood by SME, providing internal validity to the questions. The initial five-part vignette scenario developed formed a pilot test. This pilot was sent, via a gatekeeper at Dstl, to 3 SME. It is common practice for research with a small sample of SME to gain the support of a individual within the same organisation as the SME who acts as a gatekeeper and facilitates access to the SME. Following feedback from this test, the vignette scenario was refined, creating a four-part scenario that was sent out to the test group of SME (see Appendix One for complete vignette). Chapter 3 presents the phased-model of decision making specific to the air defence task that was extrapolated from the data collected through this vignette

2.3.2 Stage 2 - Questionnaire development

Through the analysis of the data collected at stage 1 of this thesis it was highlighted that our understanding of the operational use of automated systems by RN personnel remains unclear. Therefore, several questions were formulated to explore how currently serving personnel are interacting with automated systems, where these systems are used and could be used in the future, and what personnel's thoughts are on the existing procurement process of these systems (see Table 2.2). Stage 2 of this thesis is presented in full in Chapter 4 and the complete questionnaire can be found in Appendix Two.

Table 2.2: Qualitative questions included at Stage 2

1	In your opinion, do you see automated tools/systems having a role in future naval operations?
2	If so, where do you see such tools/systems having the most benefit and why?
3	During your time spent at sea or during training, how often did you interact with and utilise automated tools or systems? (i.e. daily, weekly, monthly, once etc.)
4	What were the tools you used and how did they aid your operations?
5	Have you ever been consulted in the development of new tools/systems prior to their release into operational use?
6	What are your views on the consultation of current and future personnel during the design and development stages of new automated tools/systems?

2.3.3 Summary

This chapter has so far discussed Stages 1 and 2 of the mixed-methods framework developed; the adaptation of the CDM interview into a vignette scenario questionnaire, and the development of the second qualitative questionnaire sent to SME. The qualitative data collected at stage 1 provided the narrative of the air defence task which underpinned the development of the microworld- Automatic Radar Classification Simulation (ARCS). The next section to this chapter will present Stage 3 of the mixed-method framework, which involved designing, developing and testing a micro-world as well as discussing the utility of microworlds for exploring complex decision-making environments. The initial stages to the development of ARCS will be presented before discussing the research hypotheses answered in this thesis.

2.4 Stage 3 - Microworld Design and Development

Defined as *"a computer-generated environment that exists in laboratories. It is a simplified, idealized model that adequately simulates the essential elements of a real-world system."* (Chen and Bell, 2016, p. 187). The use of simulated microworlds to explore human decision making and interaction with tools is a longstanding and continually developing research methodology. Arguably microworld application has arisen in part due to laboratory decision making tasks lacking important real-world influences (Alison et al., 2013). Additionally, when researching complex decision-making environments (characterised by time pressure, uncertainty and high stakes), simplistic experimental tasks do not enable the exploration of the complex myriad of factors that influence decision making in such environments. Microworlds embody crucial characteristics of the reality under inquiry, enabling both observation and theory testing (Eyre et al., 2008b); bridging the gap between field research and controlled laboratory studies (Brehmer and Dörner, 1993). These characteristics include:

1. **Complexity** - a goal structure is present
2. **Dynamicity** - operating in real time
3. **Opaqueness** - aspects of the system are unknown to the participant; they must make inferences and develop hypotheses on their previous experiences (Alison et al., 2013; Brehmer and Dörner, 1993; Chen and Bell, 2016).

The design of an ecologically valid microworld requires domain knowledge of the area under study. Although there is always some loss of ecological validity with abstraction of reality to a virtual environment (Loomis et al., 1999), possessing domain knowledge enables cognitive and situational demands to be woven into the system, resulting in a microworld that embodies the vital characteristics of the real-world under observation

(Gonzalez et al., 2005). The fidelity, i.e. the level of similarity between the simulation and the real world (Alison et al., 2013, p. 257) of the microworld can be achieved through interaction with SME. Further, the psychological fidelity of the simulation enables the researcher to explore the underlying processes relevant to decision making performance (Kozlowski and DeShon, 2004). With the need to ensure that new decision support systems are robust and not just applicable to a single controlled scenario (Woods, 2016) utilising high fidelity microworlds can provide researchers the opportunity to test systems, specifically human interaction with such systems, in psychologically and environmentally authentic environments. Additionally, it is increasingly important for academia to develop simulated environments that can remain open architectures, which foster collaboration as opposed to the closed systems developed by industries. Thus, the remainder of this chapter will detail the development of the micro-world the researcher of this thesis designed and had developed.

2.5 Automatic Radar Classification Simulation (ARCS)

ARCS has been designed to explore human interaction with automated decision support systems, specifically relating to the maritime above water warfare domain. The phased-decision model of the air defence task (Chapter 3) extrapolated from qualitative research conducted with SME provided the foundations for the development of ARCS. Identified from this model was the fundamental, yet often overlooked, initial stage of the air defence task- development of the Recognised Air Picture (RAP). All other actions (e.g. Threat Evaluation and Weapon Allocation (TEWA)) and decisions made are based on the RAP that is produced through this initial development of situational awareness. The higher up the chain of command you travel the pressures faced by decision makers (e.g. temporal, reputational, criticality of decision outcomes) increase. Therefore, if the initial stages of the air defence task can be expedited the pressures facing decision makers can be somewhat elevated. Further, the application of automated support tools during the initial stages of the air defence task would fit with the literature that shows the beneficial effects on performance utilising automated systems during data collection and analysis phases, as opposed to operating with automated tools for more complex cognitive tasks (such as action implementation).

The process of developing ARCS involved 4 steps; (i) building knowledge of the specifics of the research domain, (ii) designing the prototype interface and developing task structure, (iii) piloting the system, and (iv) recruiting participants and utilising ARCS

Building knowledge of the specifics of the research domain. Highlighted by (Crandall et al., 2006) as a key requirement prior to conducting CDM is the researcher holding detailed knowledge around the research domain under exploration. The same key requirement holds true for the development of micro-worlds. Terminology, work processes and command structures must all be understood by the researcher in order to enable the creation of a microworld that truly incorporates real world influences. Prior to approaching the development of ARCS, the researcher of this thesis immersed themselves in the research domain. Primarily through attending meetings with SME and maintaining informal interpersonal links with experts. Additionally, the researcher was able to observe training events during which simulations of the air defence task were observed. Additionally, the results of stage 1 of this thesis provided valuable insight into the decision-making process that underpins the air defence task, further building the knowledge based used to develop ARCS. Through developing this quasi-expert status the researcher was able to build the three key characteristics of a microworld into ARCS: *complexity, dynamicity* and *opaqueness.*

Designing the prototype interface and developing the task structure. The key benefit to utilising microworlds is to explore real world influences in a more controlled way compared to field research. The key objective was to develop a psychologically immersive environment that replicates the real-world characteristics of the air defence task. To achieve this the researcher utilised her knowledge on the research domain to design an interface that replicates, as far as was possible, the combat management system RN personnel use. The design drawings created by the author of this thesis were given to a computer programmer who then generated the java based microworld. In terms of the task structure itself, as the experiment would be run initially on university students the air defence task was simplified. However, to maintain ecological validity the key characteristics of the task, e.g. time pressure, mental and attentional demand and decisional uncertainty were built into the task. Liebhaber and Feher (2002) identified six critical cues air defence operators used to inform their classification decisions; Origin of track, IFF mode, Intelligence Report, Altitude, Proximity to airlane and EMS (Radar signature). For the purpose of this thesis and due to classification restrictions, it was not possible to build all the critical cues identified by Liebhaber and Feher (2002) into ARCS. Instead, publicly available data was utilised providing participants with the Altitude, Velocity, Proximity to airlane, Platform type and Baring as the information points from which to make their classification decisions. However, the microworld itself is flexible enough to include additional data sources and points if required at a later date. Figure 2.3 depicts the initial interface design drawing for ARCS. All information relating to the tracks on screen appears in the top left box, if a track is selected this information also shows on the top right hand side. If any track appears to be hostile, the participant is required to click the inform command button, the track information will then appear in the threat list box.

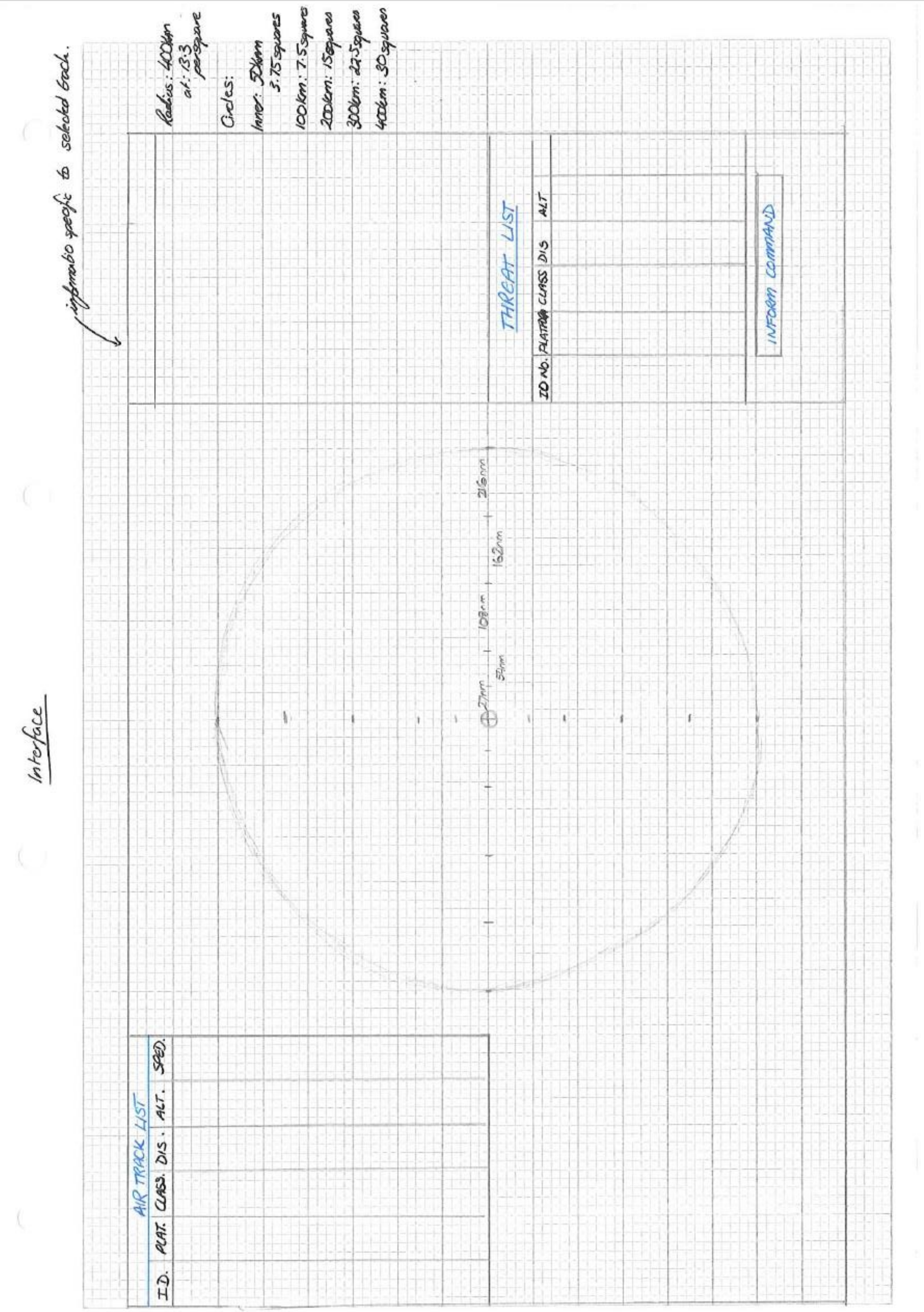


Figure 2.3: ARCS initial interface design drawing

Piloting the system. Psychological research commonly conducts pilot tests to conduct manipulation checks. For the purpose of this thesis two pilot tests were conducted to check the utility of ARCS. An overview of these pilot tests can be found in Chapter 5. To summarise the findings, the first pilot highlighted that ARCS was not manipulating workload sufficiently. Therefore, several alterations were made prior to pilot 2 which showed the utility of ARCS as a tool to explore automation usage decisions. For full details on the two pilot tests refer to Chapter 5 of this thesis.

Recruiting participants and utilising ARCS. ARCS has been designed to run two scenarios; a training scenario and the main task scenario. In total both scenarios take 45 minutes to run. However, these can be adapted and changed to suit the research question(s) under exploration. For this thesis, a practice scenario of 15 minutes and a single main task scenario of 30 minutes were designed to recreate features of a naval working environment but also bearing in mind the practicality of running long experiments with student samples. Most naval personnel involved in the air defence task will be used to working long shifts, being constantly vigilant and at their station for between 4-6 hours at a time. These time frames unfortunately couldn't be applied to the study presented in this thesis. However, ARCS can be programmed to run for any length of time, therefore future research could take up the baton and explore the effects of boredom or monotony upon automation usage decisions. Participants for the studies presented in this thesis were volunteers responding to adverts placed around the University of Liverpool campus, taking part in a repeated measures design experiment in exchange for course credit or financial reimbursement (further details on sample recruitment, the running of ARCS and the results from the experiment can be found in Chapter 6).

2.6 Questionnaire battery designed to explore individual differences and automation usage decisions

Chapter 1 highlighted the limited research into the potential influence individual differences may have upon automation usage decisions. One aspect of this thesis aims to contribute to this limited body of literature by exploring the possible relationship between deciding to use a generic automated system and scores on several cognitive trait scales. Therefore, to complement ARCS a questionnaire battery was developed including several individual differences and cognitive trait questionnaires (see Table 6.4). Morse and Niehaus (2009) argue that supplemental components to mixed methods research enhance *"description, understanding or explanation of the phenomenon under investigation"* (p19). Section B of this questionnaire consists of several qualitative questions designed to elicit rationales to contextualise system use and task performance. There has however been criticism of arguing the use of mixed methods by including a single qualitative question to a quantitative questionnaire battery (Morse and Niehaus, 2009). The questionnaire developed for this thesis was designed to include both quantitative and qualitative components that complement the microworld experimental task. This results in a questionnaire that functions as a supplemental data collection method. Further information on the choice of these measures can be found in Chapter 6 and the complete questionnaire can be found in Appendix Five.

Table 2.3: Overview of the sections included in the questionnaire battery at Stage 2

Section A	Demographics
Section B	Qualitative feedback questions
Section C	Cognitive Flexibility Inventory (Denis & Wander Val, 2010)
Section D	Need for Closure (Kruglanski et al., 1993)
Section E	Assessment & Locomotion (Kruglanski et al., 2000)
Section F	BIS-11 (Patton, Stanford & Barratt, 1995)
Section G	Conscientiousness Scale (John & Srivastava, 1999)
Section H	Propensity to Trust Scale (Merritt et al., 2013)

2.7 Measures collected and thesis hypothesis

To continue to provide a transparent view of this thesis, Table 2.4 presents all hypotheses addressed and the measures used to examine each research question. Additionally, the table provides the reader with directions to the relevant chapters where each hypotheses, analysis and discussion relating to the findings, are explained in more detail.

Hypotheses Number	Hypotheses Level	Research Question	Measures	Chapter
1	Exploratory	- What is the current state of automated systems in use by RN personnel? - Where do personnel view system use in the future? - What is the involvement and what are the opinions of current RN personnel towards the existing procurement process?	- Qualitative questions	4
2	Primary	Individuals with access to the DSS during the task (groups B or C) will perform better (i.e. correctly classify more tracks) compared to the control group (A).	- Number & Accuracy of classification decisions made	6
3	Primary	Individuals will use Decision Support System (DSS) when task demands exceed cognitive capacity (i.e. when mental/temporal workload is high)	- Workload - Time at which DSS selected.	6
4	Secondary	Individuals in low accountability condition will select DSS (more times) compared to individuals in high accountability condition who will only select DSS once if at all.	- Condition Low ($_L$) or High ($_H$) [within subjects] - Number of times DSS is selected	6
5	Secondary	Individuals in low automation prime condition (B) will select DSS more than individuals in the high automation prime condition (C).	- Condition (B) or (C). [between subjects] - Number of times DSS is selected	6
6	Primary/ exploratory	Explore the relationship between scores on each cognitive trait and automation use	- DSS selection - Scores on each cognitive trait	6
7	Primary/ exploratory	Explore the relationship between scores on each cognitive trait and performance on classification task	- Performance (number & accuracy of classification decisions made) - Scores on each cognitive trait - Controlling for DSS use	6
8	Exploratory	Are demographics balanced across each test group (A, B & C)?	- Performance score (number and accuracy of classifications) - Age - Gender - Past experience (i.e. previous study) - Current study/employment	6
9	Exploratory	Qualitative rationales for tool use	- Qualitative questions	6

Figure 2.4: Research hypotheses for Stages 2 and 3

2.8 Chapter Summary

This chapter presents the methodologies utilised within this thesis. In particular, this thesis contributes to the literature on human-machine-interaction by presenting a mixed methods approach that primarily draws upon the NDM philosophy. A fundamental criticism of recent approaches to understanding human-machine-interaction, specifically to testing new tools themselves, is the lack of ecological validity and brittleness to standard experimental approaches (Jamieson and Skraaning, 2017; Woods, 2016). This thesis will use a combination of methods, with each method informing and contextualise the understanding of the research aims. Therefore, by using both qualitative and quantitative methodologies this thesis provides a novel way in which human-machine-interaction can be explored in operational settings. Additionally, the complexity, dynamicity and opaqueness of ARCS provides the researcher with a highly realistic and flexible tool to take an applied approach to the problem space. In summary, this thesis will employ a mixed methods approach in order to: (i) present the decision making stages of the air defence task; (ii) discuss how automated system are currently used in operational settings and where they may be used in the future; (iii) discuss the existing procurement process, highlighting the barriers to effective automation application; and (iv) develop a high-fidelity microworld to explore individuals rationales of using a generic automated system when performing a threat detection task.

Chapter 3

Understanding the Air Defence Task: A Descriptive Decision Model from Perspectives of Royal Navy personnel

3.1 Abstract

An initial stage of this thesis was to explore the decision-making process and challenges faced by Royal Navy (RN) personnel. This was to develop an understanding of where automated decision support systems may have the most benefit to operations. To begin this process, this chapter describes the decision stages of the air defence task. The Critical Decision Method interview, adapted into a vignette scenario questionnaire, was completed by seven RN personnel. The results provide a high-level qualitative descriptive model of the air defence task. Three stages were identified: Observe; Identify and Classify; and Decide and Act. These stages occur iteratively with cognitive (overload and inferring of intent), ambient (uncertainty and time pressure), and organisational (capability and Rules of Engagement) factors potentially impeding the decision-making process. Shared situational awareness attained through the development of the recognised air picture underpins all decisions made within the air defence task. Therefore, this chapter recommends focusing on understanding how automated decision support systems could assist with the development of the recognised air picture.

3.2 Introduction

This thesis focuses on the air defence task, therefore, the first research question proposed by the author is - What are the decision-making stages within the air defence task conducted by RN personnel? To address this question an exhaustive review of the literature on maritime decision making was undertaken. This illuminated several seminal research projects and papers that had previously used Naturalistic Decision Making

(NDM) methods to explore military naval decision making. Notable projects identified via this review included Holt's work in 1988 which looked at assessing the requirement for decision support tools in the UK and the Tactical Decision Making Under Stress (TADMUS) programme conducted in the US (Cannon-Bowers and Salas, 1998). Both explored and produced decision-making models of the air defence task conducted by naval personnel. The aim of this chapter is to review models of decision making specific to the air defence task, the purpose of which is twofold. Firstly, to produce a detailed understanding of the problem space allowing the researcher to develop quasi-expert status, and secondly, to highlight new areas where decision support systems could be developed and implemented to support future RN operations. Therefore, this chapter will:

(i) Introduce historical research on the air defence task - Identifying the requirement to update our understanding of the air defence task performed by RN personnel today

(ii) Present findings from RN personnel of the phases of the decision-making process; and within this present the barriers and facilitators to effective decision making

(iii) Discuss implications of findings in terms of:

  (a) The persisting complexities and barriers to effective decision making

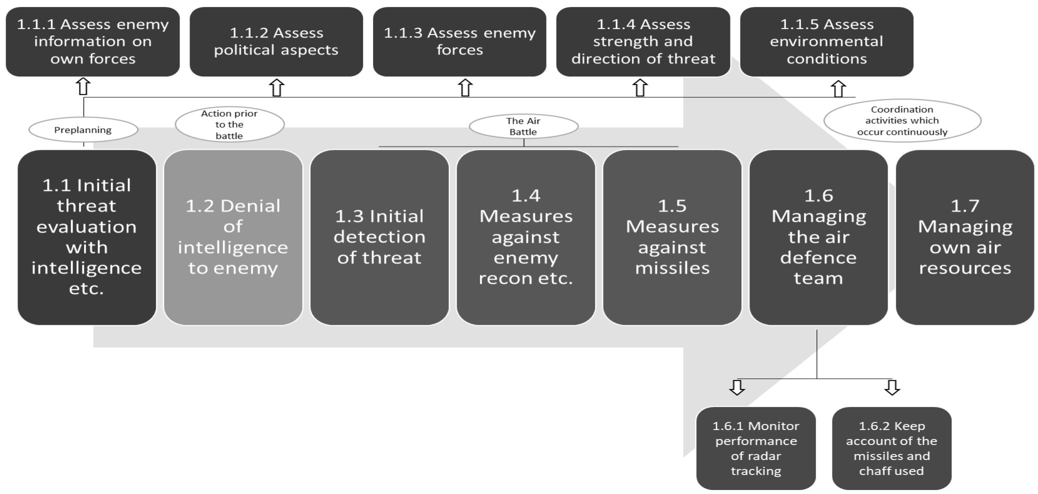
  (b) The paramount requirement to continue to extend RN capability

3.3 Historical literature on the air defence task

Teams are at the heart of all military operations (Brown, 2007), teams that are required to process vast amounts of information during both times of peace and during conflict - such as the example provided in Chapter 1. Figure 3.1 depicts the complex nature of the air defence task identified by Holt through analysis of Command and Control (C2) decisions RN personnel in Type 42 Destroyers made. Commonly, Threat Evaluation (TE) has received the most interest from academia in terms of understanding the process itself and developing tools that can support operators (Liebhaber and Feher, 2002; Roux and Van Vuuren, 2007). TE consists of identifying and prioritising objects of interest in the immediate environment, box 1.1 in Figure 3.1. TE is a fluid process with operators being required to constantly update their understanding of the current threat tracks may present (Roy et al., 2002). Due to the vast amount of information that is required in order to understand the environment and to make judgements of threat levels, it is not surprising that automated decision support systems are being developed to assist operators in this task. TE is however only one aspect of the air defence task, the stages prior to and following TE form a complex interaction of decision stages performed by all naval personnel daily. For example, TE can be followed by Weapons

Allocation (WA), the assigning of appropriate resources to counter potential threats (Helldin, 2014; Lötter and Van Vuuren, 2014).

Figure 3.1: Taken from Holt 1988, p 76 - The air defence officer's task (simplified)



Holt (1988) identified several areas where automated decision support systems would be needed for successful operations. These areas include, aiding the allocation of 'soft kill' countermeasures, Force Threat Evaluation and Weapon Allocation (TEWA), long range threat allocation, and performance monitoring of the air defence team. Soft kill countermeasures are detailed in Chapter 1 of this thesis. Force TEWA refers to the defence of several ships that are operating as a Task Group as opposed to individual units. For example, a Type 45 Destroyer (T45) may be escorting a valued asset, the air defence task then incorporates the defence of the T45 but also the asset. All TEWA decisions must be made taking into consideration the vulnerabilities and capabilities of all vessels in the Task Group, and decision support systems have been introduced to increase operational capabilities. However, it remains unclear whether the areas identified by Holt are being supported by automated systems and therefore if future research needs to focus on different aspects of the air defence task.

The TADMUS programme began in the US in 1990 following the USS Vincennes incident. On 3rd July 1988, a US cruiser shot down a commercial aircraft killing 290 people over the Arabian Gulf (Hammond, 2017). The aims of TADMUS were to increase the overall skill level of US Navy personnel through the development of stress-exposure training and decision support tools, thereby enhancing critical thinking skills, decision making and teamwork (Cannon-Bowers and Salas, 1998; Cohen et al., 1997; Driskell and Johnston, 1998; Johnston et al., 1998; Riffenburgh, 1991). The overall aim of TADMUS was to prevent errors in judgement that could result in the death of innocent civilians or lead to fratricide incidents.

The decision stages of the air defence task are practically identical across nations and take place within a highly complex military theatre (Foard, 2000). However, certain differences exist relating to Rules of Engagement (ROE) [1], number of personnel and vessels in operation, as well as available sensor technology and weapon systems. These differences will impact upon decision-making and where automated systems can be introduced. Therefore, developing up to date descriptive models of the task under exploration enables researchers to identify leverage points that can facilitate system design (Militello et al., 2015). Warfare has become increasingly ill-defined and complex, increasing inherent uncertainty with military operations in times of peace and conflict (as outlined in Chapter 1). Therefore, it is important to continually evaluate the decision stages of the air defence task to ensure that research and system development are addressing the challenges current personnel face. This current project began in 2015 and although it is anticipated that the high-level stages of the task will remain the same as identified by previous research: (i) Formation of Situation Awareness (SA); (ii) TE; (iii) WA; and (iv) re-evaluation of the situation, the facilitators and barriers to effectively achieving this task may have changed due to the increasing availability of sophisticated automated decision support systems and the changes to the nature of warfare.

3.4 Method

3.4.1 Participants

A total of N=7 Subject Matter Experts (SME) completed the online survey distributed within the UK Ministry of Defence's Defence Science and Technology Laboratory (Dstl). Participants were required to have served or be serving in the RN with experience of working within a ships operations room. Specifically, with experience of holding positions (such as Principal Warfare Officer), that would result in them being considered experts in the air defence task (see Table 3.1 for breakdown of participant demographics). Research has recommended that when conducting critical incident interviews, three - five SME usually exhausts the domain of analysis (Militello and Hutton, 1998). Although this study utilised a vignette scenario method, the questions posed to SME were based on those that would be asked if undertaking a critical incident interview. However, it was expected that the amount of data, although not the richness of this data, gained via a vignette scenario would be lower than if interviews were possible. The

[1] Rules of Engagement are the legal frameworks that define how and under what circumstances a military can use force. Each country will have their own ROEs that their state forces must adhere to.

limitations associated with this and the small sample size are therefore acknowledged in the discussion.

Table 3.1: Demographic Information of RN personnel who completed the stage 1 questionnaire

	N	Gender	Average number of years' experience	Ranks
Fully completed questionnaires	3	3x Male	12 years 1 month	1x Lieutenant, 1x Lieutenant Commander, 1x Commander
Partially completed questionnaires	4	4x Male	10 years 0.5 month	2x Lieutenant and 2x Lieutenant Commander

3.4.2 Vignette Survey

The vignette survey was developed on Qualtrics, an online survey tool. This online tool was used for ease of distribution to potential participants. Each scenario was short and specific to operations involving a Type 45 Destroyer, for example (see Table 3.2):

Table 3.2: Example of vignette scenario and questions presented to SME

Situation 1- A number of active civil airways run either through or adjacent to the assigned sector of responsibility. A daily Air Tasking Order is being issued for friendly military air traffic operating in the JOA. T45 tasked to monitor air activity over the designated hostile nation and in particular any surveillance aircraft operating over the sea.	
Q1	Explain your thought process and the information used to develop shared situational awareness
Q2	Which information is most important to your decision making process?
Q3	Please answer the following questions i List your two key priorities in response to this situation ii Please list the actions you will take to achieve each priority iii Please indicate why these actions will achieve each specific priority

Following each part of the scenario a series of open ended questions were presented. Question one was designed to elicit knowledge on what information is used to develop shared SA. With the introduction of automated systems that will be functioning as an actor in the operations team, understanding what information will need to be shared by the automated system is vital. Question two aimed to elicit key information and data points that are used by personnel to form decisions. This information will be fed directly into the development of the microworld - Automatic Radar Classification Simulation (ARCS). Finally, question 3 was designed to extrapolate how priorities are assigned, what plans are then formed to deal with these priorities and the rationales behind these decisions and actions.

Due to the novel access to SME, several follow up questions (see Table 3.3) were also included in the survey following the scenarios. These questions were adapted from the Critical Decision Method (CDM) interview methodology that has been commonly used to elicit exploratory information relating to decision making within complex environments. Additionally, the identification from personnel of the potential barriers to effective operations will aid research twofold: firstly in identifying where automated decision support systems are needed, and secondly to identify potential strategies already in use by personnel to manage uncertainty when making decisions. The survey is presented in full in Appendix One of this thesis.

Table 3.3: Example of the supplementary questions presented to SME

Q5	What do you find are the most challenging aspects of working within an operations room and why?
Q6	At times when you have experienced uncertainty with a decision/action, how did you manage that uncertainty to ensure a decision was reached?
Q7	How is information processed/presented from systems used within the operations room?

3.4.3 Procedure

The initial survey was sent to a gatekeeper, in this instance, the Dstl Military Advisor (MA) who helped develop the vignette scenarios. The gatekeeper was employed to distribute the survey link to a test selection of SME who had experience of conducting the air defence task. The gatekeeper had direct access to such SME as well as the domain knowledge to ensure that those asked to complete the survey had the relevant practical experience. Potential participants were emailed with the study information and a link to the survey. If participants were happy to take part, they were asked to provide consent through the online survey before being presented with the vignettes. Each section to the vignette was presented individually (see Table 3.2) with each new section increasing in complexity.

SME were able to write in drop down text boxes, enabling them to elaborate on each question as they saw fit. The questionnaire was kept online for three weeks to ensure that as many SME could complete the vignettes as possible. In total 7 surveys were completed by SME of which 3 were completed in their entirety taking on average 2 hours and 22 minutes to complete. The average duration of a CDM interview is up to 2 hours and it has been argued that questionnaires can provide a comparable level of detail as an interview if participants answer conscientiously (Flanagan, 1954). Therefore, although the length of time it took for SME to complete the survey resulted in a high non-completion rate (57.14%) arguably the responses for each question provided rich qualitative data. Accordingly, all responses, whether partially or fully completed are included in the analysis.

3.4.4 Analysis

Thematic analysis has been defined as *"a method for identifying, analysing and reporting patterns (themes) within data"* (Braun and Clarke, 2006, p. 79) and can, if rigorously applied, provide an insightful method of analysis for research questions. Themes emerge inductively from the data (Aronson, 1995). This results in themes that are formed being produced by the words of practitioners, promoting the ecological validity of the research. Thematic analysis was chosen over other qualitative methods as it is a flexible analysis method that has been found to be useful for producing analyses suited to informing policy development (Braun and Clarke, 2006). Themes form a holistic picture of the collective experience of the people, groups or organisations under study (Aronson, 1995).

A crucial consideration for conducting thematic analysis is the transparency of the approach taken, which can impact upon the validity of the findings; defined as *"the credibility and accuracy of processes and outcomes associated with a research study"* (Guest et al., 2012, p. 7). Traditionally, the first step with a robust thematic analysis is transcription of the data. A questionnaire was distributed online as opposed to conducting interviews. Therefore, the process of transcription consisted of formatting each respondent's answers into a word document. Table 3.4 displays the guidelines produced by (Braun and Clarke, 2006) that were taken to ensure a robust approach to thematic analysis occurred.

QSR International's NVivo 11 software is a commonly used computer-assisted qualitative data-analysis tool (Hoover and Koerber, 2011) that facilitates the management and analysis of qualitative data (Bandara, 2006; Bazeley and Jackson, 2013). NVivo does not conduct data analysis but assists the researcher with the organisation of themes, categories and codes. This support with the management of data analysis not only aids efficiency of the analysis (Hoover and Koerber, 2011), but also enables the researcher to explore diverse ways to extrapolate meaning from the data (Bazeley and Jackson, 2013). NVivo can also support the transparency of qualitative analysis through the use of memos that can record the researchers' insights and be linked to sources (Hoover and Koerber, 2011).

Table 3.4: Guidelines taken from Braun & Clarke (2006). Using thematic analysis in psychology. P87.

Phase	Description of the process
1. Familiarising yourself with your data	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas
2. Generate initial codes	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code
3. Searching for themes	Collating codes into potential themes, gathering all data relevant to each potential theme
4. Reviewing themes	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis
5. Defining and naming themes	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme
6. Producing the report	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis

NVivo 11 facilitated all stages of analysis, from generation of initial codes to report creation. The interface is designed to support inductive qualitative research by making the process of coding, theme creation and revision of themes more intuitive. Firstly, all the RN SME questionnaire responses were formatted into word documents. These documents were then imported into NVivo 11. This enabled the 'select and drag' feature to coding themes, or nodes as they are labelled in NVivo. Themes have been defined as *"units derived from patterns"* (Aronson, 1995, p. 4), consisting of recurring statements, conversation topics, expressions of feelings or opinions. Alone these units may seem meaningless however when combined within a theme it is possible to view a comprehensive picture of the area under exploration. Following the initial generation of themes, the data was re-read to review each theme against the coded items and the entire data set, making the analysis iterative in nature. Once the themes were generated from the data and reviewed, existing literature and documentation available to the author was interwoven into the findings to facilitate comprehension of the results.

The factors identified from the analysis are presented in order of occurrence within the air defence task. Quotes from domain experts are presented alongside each factor to allow the reader to judge the veracity of analysis and to provide contextual understanding into the domain area.

3.5 Results

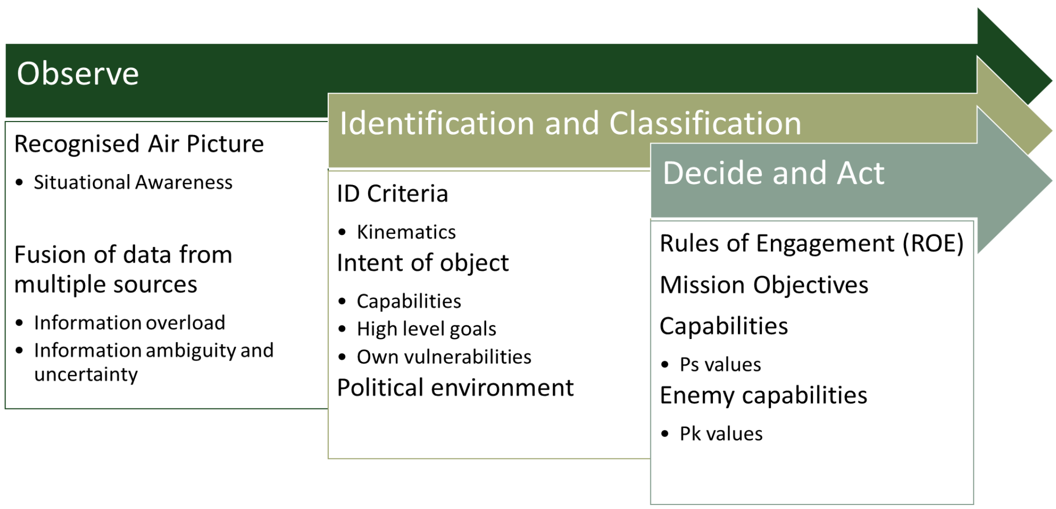


Figure 3.2: Overview of the three high-level stages of the Air Defence Task and the factors that influence decision-making derived from the data analysis

3.5.1 Air defence task

Analysis of SME responses revealed three key stages and the factors (cognitive, ambient and organisational) that impact upon each stage of the air defence task (depicted in Figure 3.2). Each of these stages will now be discussed in more detail.

3.5.2 Observe

Table 3.2 depicts the decisional requirements for the first stage of the air defence task - Observe. This stage is principally concerned with the fusion of data from the available sensors (e.g. radar, sonar etc.) and feeds directly into Stage 2 - Identify and Classify. It is only possible for Stage 2 to occur, if aircraft or objects (airbases etc.) are detected within the vicinity of the ship/task group.

Table 3.5: Decision requirements table for Stage 1 of the Air Defence Task (ADT)-Observe (adapted from template provided in Militello and Klein (2013))

Decision requirement	**Why difficult / procedural errors**	**Critical cues**	**Expert strategies**
Classify tracks; develop RAP	▪ Information avaliability ▪ Limitations of information sources ▪ Information overload ▪ Information ambiguity	▪ Civilian traffic patterns (e.g. location of airlanes) ▪ Locations of civilian and military aerodromes ▪ Daily missions flown in vicinity	▪ *"What will be the daily missions flows…including locations of friendly airbases and any patrol areas or air corridors"* ▪ *"Put in requests to discover where aircraft have been forward deployed to research aircraft flying patterns"* ▪ *"The [information overlay] is fine for traditional information (radar, ESM etc) but poor for the increasingly common forms of written intelligence"*

Fusion of data from multiple sources to build the Recognised Air Picture (RAP)

Operators are required to understand data from a number of sources (e.g. commercial flight profiles, radar, and track characteristics) pushing their cognitive capacity to its limits:

> *Range of factors are considered that may be important; track point of origin, flight profile (e.g. heading, speed, altitude, rate of climb or descent), IFF modes and codes, associated EM emissions, non/conformance with published airlines (civ) ACMs (mil). By assessing these criteria most tracks will likely be able to be identified, and any requiring further investigation quickly flagged up.*

The groundwork behind building the Recognised Air Picture (RAP) consists of the effective fusion of these data sources (providing level 1 SA) to produce a valid air picture: *"We need to monitor all [aircraft] periodically to ensure nothing has changed"*. It is from this picture (developing levels 2 and 3 SA) that all operational and tactical decisions are made: *"Establishing [a] valid RAP requires all tracks within [the] Surveillance Area to be investigated and classified - This allows Command to take further action against potentially hostile tracks"*. Resulting in the crucial requirement that this picture is accurate:

> *"Understand own and (potential) enemy strategic and operation intent. Understand sub-tactical factors: missile capabilities etc. Understand own and enemy information gathering, management and communication abilities. Understand patterns of life and neutral interests and habits. Ensure robust RAP with shared understanding."*

However, the data made available to operators is not always perfect: *"You inevitably have too much information to deal with yet not enough good information"*, resulting in inaccurate, missing or too much information being presented. The continued reference from SME to information that 'should' be readily available highlights the potential that, at times, this information may be unavailable to or indeed be missed by operators impeding the decision-making process. This would result in the RAP developed not being accurate, which could have fatal consequences for the crew and/or an ambiguous track that is engaged.

A further complicating factor with the fusion of information is that pieces of information are often distributed among the whole air defence team (Nguyen, 2002) an element also highlighted by SME: *"Obtain additional information to back up the thought* *process. It will be in the Ops rooms somewhere but not shared!"* This raises the role of effective communication as a further element to the air defence task.

Information ambiguity and uncertainty

Uncertainty is a hugely important factor that impacts upon the decision-making processes of the Air Defence team. Vast amounts of information must be fused and verified on a constantly updating basis, yet, this information is not always complete or available to the operators. In fact, operators are dealing with all four types of uncertainty proposed by (Klein, 1998); missing information, unreliable information (i.e. low confidence in the source of that information), ambiguous information and complex information. Therefore, operators are required to build the RAP whilst dealing with and allowing for many forms of uncertainty: *"Is this an exercise or indications for readiness to attack? If the later, are we in the right place? Would we survive the attack, can we best achieve our mission in another way? How can we attack them?".*

The consequences of making an error or deciding on a course of action based on incorrect or incomplete data can be fatal, adding additional pressure (i.e. the influence of potential regret or accountability concerns) onto the decision-making process: *"Need to have a completely clear picture in case of engagement- cluttered or inaccurate picture will cloud judgement and probably result in incorrect engagements- if even needed anyway".*

When asked how to deal with uncertainty one SME wrote:

> *"Refer to supervisory role in practical- e.g. [Air Warfare Officer] or if unavailable [Commanding Officer]- in order to ensure correct course of action. If impractical (time v availability of individual) results in judgement call- weighing up relative cost/benefit of differing courses of action against Command Aim. Consider ability/cost of reversing course of action and ensure that authority is not exceeded. Unless done positively no action is the wrong answer!"*

Highlighting the importance of active decision making even when faced with uncertainty, time pressures and constraints imposed by adhering to C2.

One way in which this uncertainty has been mediated, to some extent, is by the employment of technology to aid with the assimilation of incoming data. Low level processes have been automated allowing the operator to view and make judgements on already partially fused data sources. For example, the Combat Management System (CMS), which displays digitally processed radar and sonar data. However, it should be noted that operators must still make a series of decisions to categorise tracks that appear on the CMS screen, into friendly, hostile or neutral. This categorisation is based on the available information and used to produce the RAP from which operational, tactical and strategic decisions are made.

3.5.3 Identify and Classify

> *"Does this contact pose a threat to our unit? Would investigate the contact and interrogate the contact for emitter/IFF etc. Would try and establish communications with the contact. Would use the radar for height and tracking - All this information would be passed to the task force."*

Table 3.6: Decision requirements table for Stage 2 of the ADT- Identify and Classify

Decision requirement	Why difficult/ procedural errors	Critical cues	Expert strategies
Classification of tracks	▪ Information overload ▪ Information ambiguity ▪ Time criticality ▪ Communications ▪ Limitations of information sources ▪ Too many displays *"clutters your workspace and limits your ability to monitor the main console as you are turn[ing] to your right or left working on a seperate laptop"* from additional standalone equipment ▪ Inferring intent	▪ Civilian traffic patterns ▪ Locations of civilian and military aerodromes ▪ Flight profile and behaviour (e.g. altitude, speed, heading etc.)	▪ *"weighing up relative cost/benefit of differing courses of action against Command Aim"* ▪ *"training (reflexes and pattern recognition) and thought (preconceptions, analysis)"* ▪ *"rely on my training and understanding of the threat posed to my unit and act accordingly"* ▪ *"an element of gut feeling which comes with time and experience."* ▪ *"clear up the picture"*

Table 3.6 depicts the decisional requirements for Stage 2 - Identify and Classify.

Closely linked to the capability built from Stage 1, the information available to operators must now be understood in order to develop levels 2 and 3 SA; comprehension of the information and understanding of potential future states of the environment (Endsley, 2015b). The development of the RAP, a task that overlaps Stages 1 and 2 of the air defence task shows the gradual and iterative attainment of SA:

"Speed, heading and lack of friend indicators (e.g. incorrect modes/codes) are likely to be most immediate indicators of suspect activity and will be correlated with behavioural information"

"Difficult to discern intent without good understanding of local patterns of life- is this a known en[emy] tactic or is it likely to be a civ light aircraft?"

Intent

It is not only a fusion of information that builds the RAP but also the identification of the intent of each track observed within the surrounding area. How intent is inferred has long been a topic of research. Mathematically, intent can be inferred from past actions and future goals. Roy et al. (2002) argue that intent is formed of interest/desires, capabilities, vulnerabilities and opportunities. These factors must be evaluated in order to establish the strengths and weaknesses of a plan, with the rational that the intent of an object will be in accordance with a plan that holds more advantages than disadvantages. For instance, the risks of a certain course of action are lower or acceptable in relation to the potential gains from such action. Information available to operators, such as flight profile and behaviour are used to aid the inferring of intent. These theories are based on the assumption that our actions are always rational, however, the increasing complexity of warfare, including the use of tactics designed to mislead the opposing force (such as spoofing), may result in plans and intentions not adhering to rational schools of thought. As one SME highlighted: *"Depending on enemy potential missile load, climbing to 5000 feet may be higher than required for missile launch, making them suddenly less threatening. Equally, if they have missiles needing to be launched from that height, then [the] threat remains"*.

Therefore, inferring intent within this domain remains a complex: *"Low/slow flyer present difficult challenge- difficult to discern intent without good understanding of local patterns of life- is this a known en[emy] tactic or is it likely to be a civ[ilian] light aircraft?"*; flexible *"Indication of enemy intent, derived from their movements and wider intelligence."* and intuitive *"What is the [aircrafts] intention? Is it coming to us or* *something in our vicinity?"* cognitive process that requires further exploration.

Intent of the operator's own ship is also a factor during this process. Although, not providing a source of uncertainty, the intent of the ship itself, i.e. its mission objectives, must also be taken into account when conducting the air defence task. Mission objectives link into the current operating ROE and will determine what possible actions are available to deal with the current situation, and therefore must be incorporated into the engagement plans under operation.

Expertise

One possible factor that can facilitate the process of inferring intent is the application of experience and expertise: *"There is also an element of gut feeling which comes with time and experience. Training, training, training but be [sic] able to think on ones feet."* NDM research has shown how experience provides practitioners with a knowledge base from which to draw from when experiencing uncertainty, for example the use of pattern matching identified by the Recognition Primed Decision making (RPD) model (Klein et al., 2010). This knowledge base provides resilience (Hoffman et al., 2013) and also facilitates the quick recognition of potential problems occurring (e.g. the rapid identification of a suspicious track). It is this knowledge base that arguably provides the operator with the cognitive flexibility required to notice when a track is 'acting out of the ordinary' and support shared cognition to enable team co-ordination without explicit communication (Entin and Serfaty, 1999):
"Profile would be coherent with an ASM launch profile, would require consideration against en[emy] capabilities and tactics e.g. weapon range, launch profiles etc."

"Turning away at 15 miles may indicate missile release achieved. I would be very interested in the predicted performance of my radar so I knew whether or not I would have detected missile release"

Previous research found that US naval personnel were using recognitional strategies to aid their decision making in 95% of cases (Kaempf et al., 1996). Further supporting the role of expertise in facilitating effective decision making in complex environments. Additionally, training and expertise can also mitigate against the derailment that can occur with decision inerta:

"You react based on training (reflexes and pattern recognition) and thought (preconceptions, analysis), on the basis that any decision is better than no decision."

This ability to appropriately direct attention and resources could be the difference between noticing and neutralising a threat, preventing a fratricide incident or being hit by the enemy. However, due to the reduction of first hand combat experience newly qualified operators are exposed to, greater attention must be directed towards exploring novel ways in which to provide operators with the benefits experience can provide. For instance, the benefits of training in immersive simulation environments or the development of tools that can aid with directing attention to certain data points or with scheduling tasks.

3.5.4 Decide and Act

Table 3.7 depicts the decision requirements for the final stage of the air defence task- Decide and Act:

> *Ok, things are hotting up now but still need to act within ROE. Need to ensure force disposition is suitable to match threat activity but do not necessarily want to give away too much of own activity- Consider force movement to reduce likelihood of simultaneous attack - give all units time to take each threat in turn rather than fight all at once. Remain conscious of mission and political task. Review force sector assignments and ensure protection provided to high value assets.*

Capability

Throughout the task operators must remain aware of their ships' capability whilst simultaneously continually updating the RAP and their own (and team) situational awareness: *"Have weapons been fired at me?... If weapons are fired they may not be detected until at very close range, preparations should assume they have been fired and be started at the earliest engagement opportunity"*.

This capability will feed into the potential decisions available to the team depending on the type of threat faced, for example using a helicopter as opposed to a fixed wing aircraft to make contact with an unidentified track. However the SME also highlights their awareness of the limited engagement capability this would leave them with should the helicopter need to take such actions: *"Difficult to achieve intercept with friendly aircraft due to profile however should be attempted. Use of helicopter may be more practical (although limited engagement options)"*.

This shows that, in addition to their own capability, that of the enemy's must also be taken into consideration alongside any actions decided upon (e.g. what weapon

Table 3.7: Decision requirements table for Stage 3 of the ADT- Decide and Act

Decision requirement	Why difficult / procedural errors	Critical cues	Expert strategies
Classification of tracks	▪ ROE considerations ▪ Own capabilities (soft or hard kill options) ▪ Enemy capabilities (if known) ▪ Pk and Ps values	▪ Political Policy Indicator ▪ Mission Objectives ▪ Boundaries of territorial waters /airspace ▪ Communication with Task Group Commander/Headquarters	▪ *"Potentially call aircraft on radio if permitted"* ▪ *"Robust application of ROE will require good understanding of force disposition- e.g. which airfract are weapon carriers, what range weapons can be launched, what targeting is required for different weapon systems"* ▪ *"Robust comms and passing of SA to other units key, need to ensure they feel confident T45 can conduct rapid engagement without risk to force if necessary"*

system may the enemy be carrying, what range, number of missiles etc.). The Probability of Success (Ps) and Probability of Kill (Pk) values have long been referred to within the literature in relation to optimising automated threat assessment systems (Bard, 1981). Each engagement plan must take into consideration these variables in order to develop the plan which allows the greatest likelihood of mission success and probability of escaping harm. When facing multiple threats, this probability may not equal 100%, therefore the Command Team are, again, acting under uncertainty with the added stress of facing a potentially fatal situation. Understanding the enemy's and their own capabilities is vital to help reduce some of this uncertainty as is the requirement for operators to be cognitively flexible: *"[As] full an understanding as possible of en[emy] forces and disposition, e.g. number of units and weapons carried- allocation of sufficient weapons to defend against threat".*

Constant re-evaluation of the situation is still required once an action is taken in order to evaluate that action and re-form a plan based upon that evaluation. If an action fails the first time (e.g. the soft-kill action failed to deceive the enemy or the missile misses the target) the team must decide whether to fire again, to try a different manoeuver or to double check that the track is indeed an enemy missile or aircraft: *"Minimise weapon expenditure- need to be here a while? Ensures safety of own and civil aircraft".*

Rules of Engagement (ROE)

Although some low-level processes have been successfully automated, facilitating decisions made within the air defence task and the operators' daily tasks, there remains reticence to move automated functions up the chain of command. This reticence is potentially due to accountability concerns and the ability to ensure ROE are satisfied and adhered too:

"To determine if an engagement may be warranted in order to defend the force. ROE may require positive (visual) ID, additionally friendly aircraft may deter aggressors. ROE owner will require full narrative to enable decisions/delegation of ROE as required."

"Robust application of ROE will require good understanding of force disposition- e.g. which aircraft are weapon carriers, what range weapons can be launched, what targeting is required for different weapon systems".

These concerns may also underpin the reticence of the operators themselves to utilise the systems they currently have available to them. However, there is also an opposing drive to continue to utilise automated systems due to the benefit and support they can, when utilised effectively, provide the operator.

> *Better use could be made of automated systems to apply e.g. ID CRIT, with machine learning used to enhance accuracy. E.g. rather than classifying tracks manually, the system could suggest IDs for the operator to approve or veto and the system could refine its algorithms based on this feedback. Similar automation could be applied to identifying potentially suspect tracks, Would require operator supervisions due to ROE issues.*

It should be noted that no responses suggested a move towards complete automation at any stage of the air defence task. Arguably, such a complex task in which ROE are paramount, will always require a human element.

3.6 Discussion

This research presents detailed accounts from domain experts of the air defence task conducted by RN personnel in Type 45 Destroyers operating in littoral environments. As the spectrum of threats and the range of enemy tactics increase in complexity, the capability of our forces must continue to be extended, arguably through the synergistic collaboration between man and machine. The principles of NDM have been applied with this research to elicit detailed understanding of the current cognitive, ambient and organisational factors behind decision making during the air defence task. A descriptive phased model of decision making is presented providing the groundwork to explore the potential for new automated systems to be adopted.

The three stages of the air defence task are presented: Observe; Identify and Classify; and Decide and Act. It should however, be noted that the phases are not linear but iterative, with each phase building upon the previous. Conceptually, the task is cyclical in nature, with operations occurring at each stage simultaneously (see Figure 3.3), mirroring the SAFE-T model of decision making proposed by van den Heuvel et al. (2012). For instance, building the RAP is a continual process that occurs as SA (Situation Assessment phase) is passed throughout the crew and up the chain of command. Command will then make decisions (Plan Formation phase, F) and order actions (Plan Execution phase, E) based on the continually updating the RAP. Following each decision and action the cycle returns to the beginning, re-evaluating the new current state of play (Team Learning phase, T).

Derailments from this decision process occur, primarily due to the sheer volume of information that operators must process and which places RN personnel under extreme cognitive load. The difficulty of a task is driven by the volume of information needed to be processed, not necessarily the complexity of that information (Sweller et al., 2011). However, air defence operators are not only dealing with large volumes of information, but also with the complexity of that information (i.e. the ambiguous nature of the available information and uncertainty surrounding inferring intent). Nevertheless, it should be noted that RN operators have been performing this complex task for decades and the ability of a well-trained and experienced operator to perform this task is not in question. What is of concern is that with the increasingly complex threats that our military face in terms of new enemy tactics and weaponry, the amount of information present during theatre is rapidly increasing and at times moves beyond the ability of any human to process in the required time available (Benaskeur et al., 2008).

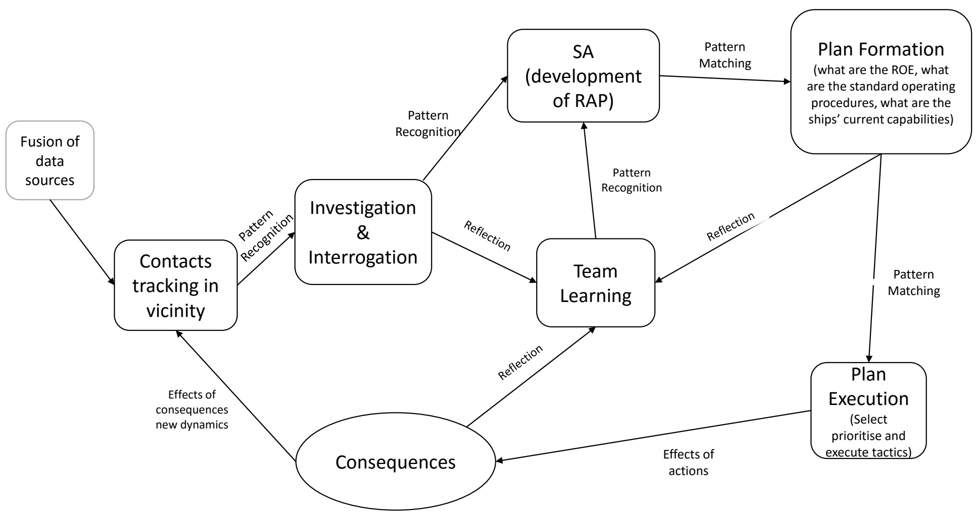


Figure 3.3: SAFE-T Model of decision making adapted to the air defence task (basic model structure from van den Heuval, Alison & Crego, 2012)

Similarly, to the findings of Kaempf et al. (1996), the research described here indicates that operators are working within constantly shifting environments, requiring the updating of SA, achieved primarily through gathering and sensemaking of information available within the operations room. Arguably, the main task conducted within air defence operations is the building of the RAP. Decisions made beyond this stage (e.g. should a threat be engaged? how should such engagement take place?) are procedural in nature since ROE and standard operating procedures must be followed. Personnel highlighted the importance of transparency of decision making rationales in order for ROE to be complied with, systems that play a role in decisions that must adhere to ROE will therefore need to function in a transparent manner.

Do Holt's recommendations remain relevant today?

As mentioned in the introduction Holt identified four areas that would benefit from the application of automated decision support systems. These areas include, aiding the allocation of 'soft kill' countermeasures, force TEWA, long range threat allocation and performance monitoring of the air defence team. The responses from SME presented in this chapter highlight that these areas are still of importance for modern naval operations. The formulation of engagement plans must take into consideration the use of 'soft kill' countermeasures and 'hard kill' weapons, for instance, how many resources are on the ship and what is the knock-on impact of using a countermeasure. When travelling as a Force, these engagement plans must account for the movements and capabilities of all vessels in that force, for instance, would releasing a 'soft-kill' countermeasure to protect the asset move the threat onto another vessel? Communication between individuals and teams (at Force level) can if effective facilitate the development of SA or if ineffective become a barrier to the development of distributed SA thereby impacting upon decisions made. Research has been focusing on how automated systems can support these areas, for example CORALS (Benaskeur et al., 2011). However, how operators interact with and how these systems support personnel in operational environments requires further attention from research.

Limitations

Due to the limited access to SME it was decided to adapt the CDM interview to a questionnaire format. Interview methods are subjective, and although there is a semi-structured plan with CDM, no two interviews will be exactly the same. Therefore, adapting the CDM interview into a vignette survey produced a standardised data collection method for this study. The application of an online vignette survey also removed the influence of interviewer bias, therefore all responses were driven from participants and were not influenced by having an interviewer present. It is however acknowledged that survey responses may not elicit as much detailed information as an interview would, nor would allow for elaboration on any salient points that are raised within responses. Nonetheless, as the aim of this study was to explore and elicit knowledge from domain experts (and with restricted access to such experts) it was deemed that an online survey was a suitable method from which to gain valuable insight from SME. NDM applications enable depth of understanding from small samples. However, the small sample of this work is acknowledged as a limitation. This limitation is somewhat mitigated by the level of detail provided by each expert providing a small, yet rich, data set. The limited availability of experts in this field was a further barrier to collecting a larger sample size for this study.

Although rich qualitative data was gained from SME it is acknowledged that questionnaire responses may not completely reflect operations when personal are at sea, and critically within hostile environments. It may be that, like van den Heuvel et al. (2012) discovered, in practice the decision phases do not occur in a linear cycle, stages are sometimes skipped or those stages occur in a differing order. It is therefore proposed that this study provides a starting point from which further research (utilising observational and live simulation data) can build upon. Future research should also explore if the phase-decision model presented here holds true during both times of peace and of war.

Finally, with hindsight it is acknowledged that thematic analysis may not have been the most appropriate method of analysis for this study. As the findings strongly support prior models that have been developed to depict the air defence task and due to the author having an awareness of these models during the data collection and analysis, it is not possible to pull apart if the analysis was discovering an artefact of previous literature or a truism. With all research there is a balance that must be struck between using the most appropriate analysis method for the data whilst also considering the experience and skills the researcher has. As discussed in Chapter 2 utilising mixed-methods requires the researcher to draw on prior experience with certain methods to reduce the potential impact lack of skill with certain forms of method or analysis may have. The author of this thesis had primarily used thematic analysis in prior research and applied this analysis method to this study, however if this study were to be conducted again it is suggested that an alternative data analysis method be used, for example content analysis.

3.6.1 Conclusions

This chapter sort to explore the decision stages of the air defence task that serving personnel experienced. Presented is a detailed overview of the air defence task, updating historic research and confirming that the high-level stages of the air defence task (Observe, Identify and Classify, Decide and Act) remain. Naval personnel are still operating in highly complicated, time pressured and critical environments. The introduction of automated systems at low levels, for example personnel no longer observing raw radar and sonar data, has reduced some of the technical pressure of the task but arguably has not reduced the volume of information that personnel must process as individuals and teams in order to make effective air defence decisions. The findings from this chapter fed directly into the development of ARCS, providing the researcher with domain specific knowledge that was crucial in informing the design of ARCS to be ecologically valid.

Principally, the same areas that Holt identified in 1988 that could benefit from automation remain the same. Which raises questions on the focus and impact of automated decision support development and procurement. Is it that systems have been developed and are not used appropriately, thereby not providing the support to personnel? Or are the systems that have been developed not fit for purpose and therefore do not support personnel in these key areas? The following chapter presented in this thesis provides answers to these questions.

Chapter 4

On the bridges: insight into the current and future use of automated systems as seen by Royal Navy personnel.

4.1 Abstract

This chapter explores the current state of automated systems and how they are used in the Royal Navy (RN), as well as exploring where personnel view systems would have the most benefit to their day-to-day operations in the future. Additionally, personnel's views on the current consultation process for new systems are presented. Currently serving RN personnel (N=46) completed an online questionnaire distributed at the Maritime Warfare School. Thematic analysis was conducted on the 5125 words that were generated by personnel. Results show that RN personnel understand the requirement to utilise automated systems in order to maintain capability in the increasingly complex environments they face. This requirement will increase as future warfare continues to change and increasingly sophisticated threats are faced. However, it was highlighted that current consultation and procurement procedures often result in new automated systems that are not fit for purpose at time of release. This has negative consequences on operator tasks, for example by increasing workload, and reducing appropriate system use, as well as increasing financial costs associated with the new systems. It is recommended that an increase in communication and collaboration between currently serving personnel and system designers may result in preventing the release of systems that are not fit for purpose.

4.2 Introduction

This chapter builds upon the previous, presenting unique insights from currently serving RN personnel. This section explores where automated systems are currently in use by RN personnel when at sea and where current personnel view future developments to progress. Additionally, this chapter presents an understanding of the level of involvement between system designers and current personnel, answering the question of if RN personnel are included in the consultation and development stages of new systems. Therefore, the aims of this chapter are as follows:

(i) Present the literature on current use of automated systems in military contexts

(ii) Outline the future visions of the RN for 2025 and 2045

(iii) Present qualitative findings on where automation is currently in use, how RN personnel view this use to evolve in future operations and the level of involvement current RN personnel have with system development

(iv) Discuss the results with reference to the literature and current doctrine, and to provide recommendations

Recently, an increase in use of automated systems within the RN has been seen (Royal Navy, nd). This is due to increasing operational complexity across all military environments (Benaskeur et al., 2011), combined with a decrease in available manpower. A wealth of literature already exists exploring the ways in which automated systems can increase operational capabilities for example see, Dzindolet et al. (2001); Röttger et al. (2009); St. John et al. (2005). However, in their review on the social, cognitive and motivational factors that influence automation reliance, Dzindolet et al. (2010) found that highly reliable automation does not always lead to performance improvements. This suggests that socio-technical systems, where humans and machines work collaboratively (Hoffman and Militello, 2012), are highly complex and require researchers to explore multiple factors that underpin the efficiency of such systems.

In the July Task Force Report published by the Department of Defence (Murphy and Shields, 2012), automated systems are argued to extend and complement human capability, not to replace it. Similarly, one vital aspect of RN Rules of Engagement (ROE) is that a human remains *in* the decision-making loop. Therefore, personnel across all military platforms are having to adapt their roles to become supervisors to automated systems. Research has shown the derailments that can occur with this job role transition; namely loss of situational awareness, 'out-of-the-loop' phenomena (Endsley and Kiris, 1995) and loss of manual skill (Casner et al., 2014). The introduction of adaptive automated systems is posited to prevent these derailments from occurring (Parasuraman and Wickens, 2008). However, of vital importance for the development of appropriate adaptive automated systems is to fully understand the job role or decision process that the automated system will become a part of. The Tactical Decision Making Under Stress (TADMUS) programme is one example of a comprehensive research project that aimed to develop decision making aids for low-intensity conflict (Cannon-Bowers and Salas, 1998; Cohen et al., 1997; Driskell and Johnston, 1998; Johnston et al., 1998; Riffenburgh, 1991). Decision centred design has also provided researchers with a repository of methods that can be used to elicit knowledge and domain understanding from expert practitioners (Crandall et al., 2006). This knowledge can then be leveraged to develop adaptive and effective automated systems that operate alongside their human counterparts. As Militello and Klein (2013) argue *"design approaches that are insensitive to expertise run the risk of creating designs that interfere with the development and application of skills"* (p261).

4.2.1 Current doctrine around automated systems

The Navy is a versatile force, reflected in the domains the RN is required to operate in: above water, underwater, mine countermeasures and land and littoral manoeuvres. There are three core roles that the RN fulfils, war-fighting, maritime security and international engagement (Ministry of Defence, 2011). War-fighting consists of being ready for engagement at sea or from the sea, maritime security encompasses protection of the lawful and safe use of the sea where UK prosperity and security is concerned, and international engagement relates to the promotion of UK interests through the development of international partnerships. Open source British Maritime Doctrine (Ministry of Defence, 2011) does not specifically mention the use of automated systems to support the RN in performing its functions. However, looking specifically at decision making within the UK military, caution is voiced about the failure of automated systems and that automation cannot make decisions based on intuition and empathy. The doctrine cites that, *"Human involvement in the analysis process is thus an enduring requirement"* (Ministry of Defence, 2010, p. 1-5). The 5th edition of British Maritime Doctrine has been recently published and therefore with the increasing speed of technological development it is highly likely that the use of automated systems within the maritime environment is discussed within.

Good decision making is required to succeed in combat, and information superiority is necessary to making good decisions. This superiority can be achieved by timely and accurate information (Ministry of Defence, 2011), appropriate strategy development and prosecution, well-trained and calibrated teams, and facilities that are fit for purpose (Larken, 2002). Automated systems are now inherently part of the facilities that support good decision making. However, a recent review highlighted that current automated systems can provide more capabilities than operators can understand or use (Strauch, 2017). This can result in increasing operator workload if automation performs unexpectedly (Kaber et al., 2006) - coined as "automation surprises" (Sarter et al., 1997) which have been associated with decreasing acceptance of automated systems (Ghazizadeh et al., 2012; Woods, 2016). Therefore, one way in which to extend operators understanding of the functionality of automated systems is to include them in the design and development process. The Athena project in the US is a forum designed to facilitate sailors in the US Navy to put forward their ideas to improve Navy operations and Command. This form of open communication between Navy personnel, academics and industrial partners is argued to create forward thinking sailors for the "Fleet of tomorrow" ("The Athena Project", 2013). However, this organisational support for open communication is not common globally (Elgafoss et al., 2009).

4.2.2 Future visions of 2025 and 2045

Increasing use of automated and unmanned systems is occurring and will continue as technology develops and becomes cheaper. However, military decision making is proposed to remain the remit of human personnel, primarily for ethical reasons (Ministry of Defence, 2014). The financial investments required to develop new automated and unmanned systems or robots combined with the defence budget restrictions will mean that most developments will occur in commercial settings (Ministry of Defence, 2014). In the next 10 years the RN is intending to bring into service new Wildcat and Merlin helicopters, a second Queen Elizabeth class carrier and Type 26 Global combat ships. However, there are concerns around the length of time required for these new platforms to be fully operational (Newson, 2016).

Additionally, the 2015 Strategic Defence and Security Review pledged to increase the RN/Royal Marines (RM) manpower to 30,450 by 2020. At the time of this pledge the RN/RM manpower stood at 29,710 (Newson, 2016). This number has steadily declined since then, with the Ministry of Defence statistics of 1st October 2017 showing 29,280 RN/RM personnel in service (Ministry of Defence, 2017). This deficit in manpower results in the greater need for automated systems to fill the gap, working collaboratively with personnel in increasingly technical socio-technical systems.

The highest ranked country for military expenditure is the USA, in 2016 investing \$611 billion (Tian et al., 2017). This relatively sustained level of funding has allowed the United States to invest in large scale research projects, for example TADMUS, allowing close collaborative working relationships to develop between academic organisations and branches of the military. This collaboration is also seen across the aviation industry, with the Federal Aviation Authority including human factors research into their national plan (Johnston and McDonald, 2017). However, identified by Krueger (2012) is the requirement for the US to *"master the correct mix of manpower staffing and personnel skills required for development of the new generation of naval vessels* *that envision operating with significantly reduced manning rosters"* (p238). Therefore, although the RN could explore developing concepts created in the US, for example the Athena Project, the US could also learn from the RN who have been operating with limited manpower for several years.

It is therefore important to explore the current state of automated systems and how they are used in the RN, as well as to explore where personnel view systems would have the most benefit to their day-to-day operations in the future. Additionally, this chapter explores if current RN personnel are consulted in the procurement of new systems, what their views are on the current consultation process and the impact this has (if any) on their daily use of automated systems.

4.3 Method

4.3.1 Participants

Data was collected from N=46 Royal Navy personnel via a gatekeeper at Dstl seconded to the Maritime Warfare Center (MWC). MWC operates at HMS Collingwood, the Royal Navy's largest training establishment. The Maritime Warfare School (MWS) delivers training on all operational areas covered by the RN, for example in weapon engineering, sea survival and chemical biological radiation nuclear and damage control. MWS is part of the Flag Officer Sea Training (FOST) organisation and aims *"to train Officers and Ratings for the Fleet who are ready to fight and win."* (Royal Navy., nd). A range of training courses for personnel at various stages of their careers are held throughout the year at MWS. This is reflected in the sample of RN personnel who completed the questionnaire. The variation between participants in their previous job roles, their current roles and the amount of time spent at sea provides a comprehensive picture of the opinions of RN personnel at various stages of their careers. Access to this data was enabled by the author developing a close working relationship with the gatekeepers at Dstl. Following previous collaboration (see previous chapter), several meetings were held to present the planned research as well as making sure all appropriate channels were observed. For instance, the study received ethical approval from the University of Liverpool and MoDREC (Protocol No: 785/MoDREC/16). With the gatekeeper on board with the project it was then possible to get approval from the appropriate persons at MWC and MWS to ensure that the questionnaire was passed out to all relevant SMEs.

4.3.2 Materials

Prospective participants were given the questionnaire pack (Appendix Two) which contained the information sheet explaining the purpose of the study, the consent form and all sections of the questionnaire (see Table 4.1). This method of distribution allowed each participant to take 24 hours before deciding to take part in the study. To acknowledge their consent each participant was asked to sign the consent form before completing the questionnaire. Section A asked participants to provide basic descriptive details on age, sex, current job role, previous job roles and time spent at each, length of time spent at sea during their career and how long it has been since they were last at sea. Section B consisted of 6 open-ended questions designed to identify where automated or semi-automated systems are currently being used in service (questions 3 & 4), where current personnel see automated systems being used in future operations (questions 1 & 2), and their level of involvement in the procurement of new systems or tools (questions 5 & 6). Table 4.2 displays each question included in full.

Table 4.1: Sections included in the questionnaire pack for Stage 2

Section A	Demographics
Section B	Qualitative questions

Table 4.2: Qualitative questions included in RN personnel questionnaire battery in Stage 2

1	In your opinion, do you see automated tools/systems having a role in future naval operations?
2	If so, where do you see such tools/systems having the most benefit and why?
3	During your time spent at sea or during training, how often did you interact with and utilise automated tools or systems? (i.e. daily, weekly, monthly, once etc.)
4	What were the tools you used and how did they aid your operations?
5	Have you ever been consulted in the development of new tools/systems prior to their release into operational use?
6	What are your views on the consultation of current and future personnel during the design and development stages of new automated tools/systems?

4.3.3 Analysis

Thematic analysis was conducted upon the data, supported by using NVivo 11 software. Chapter 3 describes the stages of analysis in full. However, in brief all the data was read and re-read to generate initial codes. These codes were then collated into potential themes. The next stage involved reviewing these themes in relation to the coded extracts and the entire data set to judge their veracity. Each theme was then defined and named before creating the final report of the analysis.

4.4 Results

4.4.1 Participants

A total of N=46 RN Subject Matter Experts (SME) completed the questionnaire. One SME requested to withdraw from the study, leaving n=45. The majority of participants were aged between 25-34 years, ranging from 24 to 54 years. Most of the sample were males (n= 38) and two SME declined to provide their age or gender. Number of years of service ranged from 32 years to 4 years 5 months, on average SME had 11 years' experience. Time spent at sea also ranged between SME, from 480 days to 17 years. Time since last at sea on average was 19 months (excepting the two SME who had more recently been at sea, 2 days and 10 days prior to completing the questionnaire), and ranged from 2 days to 15 years. Table 4.3 presents a selection of SME current and previous job roles.

Table 4.3: Examples of SME current job roles and previously held roles within RN

Examples of Current Job Roles	Examples of Previous Job Roles
Officer in charge of Advanced Warfare Training	Engineering Officer
Warfare Officer	Officer of the Watch
Principal Warfare Officer Training Student	Navigator
Electronic Warfare Specialist	Gunnery Officer
Fighter Controller	Commanding Officer

N.B to preserve anonymity of participants the current job roles presented do not match to the previous roles presented.

4.4.2 Within subject differences

Chi-squared analysis revealed no significant associations between age, gender, time in RN, time since last at sea and the number of coded references for each theme that emerged from the data. Conceptualising the level of detail provided by each expert by the number of words included in the coded references it was possible to explore the similarities between experts that provided the most details. Taking the top five contributors to each theme, 4 of the 5 were aged between 25-34 years, with three of those four having had 10-15 years' experience in the RN and had been at sea in the last 6-12 months.

4.4.3 Questionnaire

A total of 5125 words were generated by SME for the six questions contained in Section B of the questionnaire pack. Initial line by line coding created 272 nodes. These nodes were then analysed using thematic analysis in NVivo11 to identify emerging themes.

Themes were articulated around the four research questions: 1) What are the current areas of operation that utilise automation/semi-automated systems, 2) Where do RN personnel view automated systems being used in future operational environments, 3) to explore the level of engagement between practitioners and system developers, and 4) to explore current RN personnel opinions of the consultation process. Quotes presented throughout the results are taken directly from SME responses to allow the reader to judge the veracity of the themes that emerged. Where words are underlined this reflects the emphasis placed by the SME on certain words or phrases. Where text was illegible [...] will symbolise the omitted words.

4.4.4 Results summary

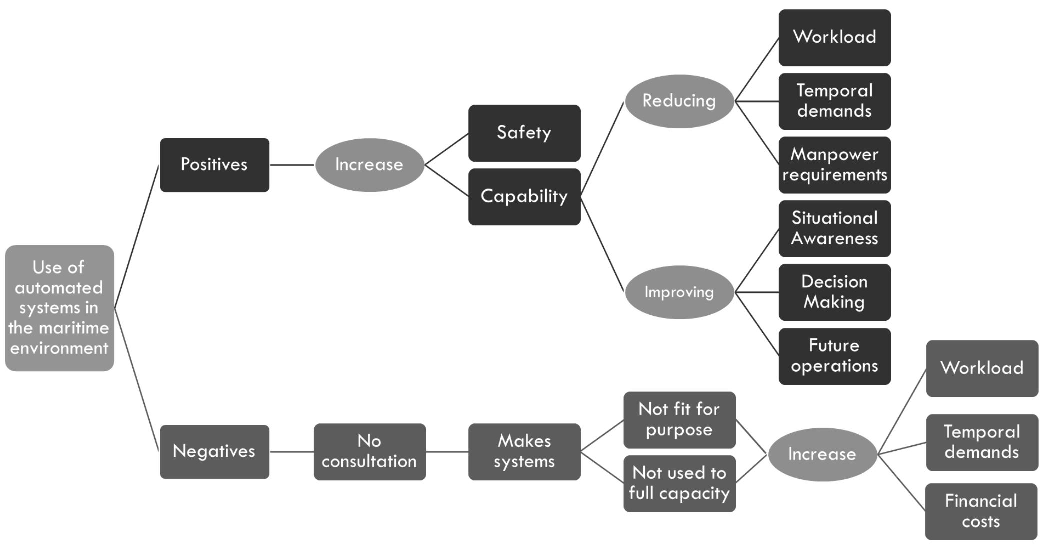


Figure 4.1: Superordinate view of current use of automated systems in RN derived from the data analysis

Figure 4.1 depicts a superordinate view of the themes that emerged from thematic analysis of the data. Themes are shown articulated around the identified positives and negatives of automated system use and how the consultation process impacts upon this. Overall SME were positive towards the use of automation citing the increase in capability that is obtained through system use. Further, currently serving RN personnel highlighted the increasing requirement to utilise automated systems across all domains of the RN. Future operations will inherently involve increasingly technical socio-technical teams. Therefore, it is of paramount importance that new systems are developed to be fit for purpose at time of release into service. However, the current lack of consultation of SME results in automated systems that are not fit for purpose at time of release. This negatively impacts upon the ability for systems to support personnel and increase capability, instead the introduction of these systems increases the workload of personnel. Additionally, further financial costs are then endured to fix the operational problems with the system. SME highlighted the values of including current serving personnel in the consultation and procurement process. Collaboration and the knowledge transfer that ensues would enable systems to be built fit for purpose at release date thus increasing capability as they are designed with the current needs of personnel in mind.

4.4.5 Inter-rater reliability

A commonly used statistical method to analyse inter-rater reliability is Cohen's kappa, which determines the level of agreement between the primary and secondary coder. The secondary coder was blind to all research questions and was given 10% of the coded references to categorise. There were eight categories that the coder could classify a quote into and a single quote could belong to multiple categories. Therefore, kappa statistic was run for each category providing a range of inter-rater reliability scores. It is commonly cited that kappa statistics of 0.00-0.20 indicate slight agreement, 0.21-0.40 indicate fair agreement, 0.41-0.60 indicate moderate agreement, 0.61-0.80 indicate substantial agreement and 0.81-1.00 indicate almost perfect agreement (Landis and Koch, 1977). Figure 4.2 presents the inter-rater reliability for each identified theme.

Theme	Subtheme	Kappa statistic	Level of agreement
Capability	Sensor	κ = .052 (95% CI, 0.13 to 0.44), $p < .793$	Slight agreement
	Personnel	κ = .182 (95% CI, -0.14 to 0.49), $p < .269$	
	Tactical decision making	κ = .505 (95% CI, 0.14 to 0.87), $p < .013$	Moderate agreement
	Weapons systems	κ = .466 (95% CI, 0.08 to 0.85), $p < .021$	
Safety	Navigation	κ = .582 (95% CI, 0.16 to 1.00), $p < .006$	
	Caution	κ = .792 (95% CI, 0.52 to 1.06), $p < .000$	Substantial agreement
Financial		κ = .773 (95% CI, 0.48 to 1.07), $p < .000$	
System Design		κ = .645 (95% CI, 0.01 to 1.28), $p < .001$	

Figure 4.2: Inter-rater reliability for each theme

4.4.6 Current operational use of automated systems

A total of 223 nodes (consisting of approximately 4,120 words) emerged from the data when exploring research questions one and two. Current use accounted for 25.6% of the total quotes identified, future use accounted for 54.3%. From the two questions discussing current tool use two themes emerged, capability (16.6% of total quotes) and safety (8.5%). Within future use four themes emerged from the data, capability (32.3%), safety (11.2%), financial (6.3%) and system design (4.5%).

It emerged from the data that automated systems are currently in operation across all aspects of the maritime operational environment, as one expert wrote automated systems are *"inherent to operating/living on a ship"*. This is in line with the current requirement to better understand how humans interact with varying levels of automated systems. SME identified currently used operational tools that increase RN capability across all four of the domains covered by RN operations; above water, underwater, mine countermeasures and land and littoral manoeuvre. Therefore, the main theme to emerge from the data related to increasing capability. For example, NAUTIS the name of the Command and Control (C2) system that controls equipment and combat systems that enable mine countermeasure missions to be conducted with increased accuracy and reduced risk to RN personnel. One such autonomous underwater vehicle mentioned by participants is Remus 100, which is a *"survey system [that] identified bottom contacts for investigation by divers removing requirement to risk ship proceeding through hazard area to conduct mine countermeasures"; "The vehicle would transit to a contact unaided by the operator saving battery time and achieving more accurate assessment of underwater contacts"*. Additionally, sensor capabilities currently support the development of situational awareness, as one expert wrote they *"improved our understanding of a theatre"*, across both underwater and above water domains, *"Automated picture compilation on radar systems, Computer aided classification system on sonars"*.

However, the data also highlighted how several of these systems are not currently used to their full capacity. Referring to the command system in use in the operations room one SME wrote that the command system *"Also has many automated response functions that are not currently used"*. Another SME wrote about the lack of trust in certain systems, *"we therefore double our workload by continuing to do everything manually as well"*. These comments suggest that the human-automation interaction that exists requires further exploration to continue to develop trust in available systems, as well as develop an understanding into how to more effectively utilise current operational systems to their full extent. It is also worth noting the reference made by SME to the continual adaptation of operators themselves when new systems are introduced. Referring to mine countermeasure operations one SME stated that *"once aware of nuancey operators evolved their ability to employ the system"*, another stated *"The art is knowing when to adjust these systems"*. Suggesting that personnel are required to continually adapt and expand their knowledge on the functioning of new systems after they have been introduced. This brings into consideration the training that personnel receive prior to their deployment and, perhaps more importantly, continued training whilst on tour, especially if a new system has been brought into service and is expected to be used operationally at sea.

A second, smaller theme that emerged from the data centres around safety, of personnel and vessel(s). Most commonly mentioned was the use of automated navigation systems, cited by 17 of the SME. Current navigation systems perform a range of functions from *"auto-pilot (heading keeping tool) on the bridge"* to *"WECDIS [1] Navigation system- auto displays ships position and reduces chartwork burden"*. Hazard warning systems play a key role in the current makeup of automated systems in use. These warning systems prevent collision of vessels, via the navigation components, and aid with the maintenance of the ship(s). For example, *"Automated damage controls system: allow for greater situational awareness when fire & floods occur"*.

Improved safety to personnel and vessel(s) arguably is a perceived benefit of adopting automated systems. Further benefits were identified by participants in their responses. For example reducing workload: *"Combat management systems- reduce operator workload in managing the tactical picture"*; aiding situation awareness: *"Autopilot on ship's helm reduces helmsman steering burden to enhance visual outlook, AIS (Automatic Identification System) aided situational awareness with information displayed on WECDIS navigation system automatically"*; coping with reduced manpower: *"Machinery controls systems- aided by reducing manpower requirements & safeguarding equipment"*; reducing time consuming tasks: *"ECPINS [2] - massive reduction in time spend fixing the ship, Radar- automated systems to automatically hook & track contacts improves reaction time, especially in the air environment"*; and supporting effective decision making: *"Automated ID of tracks within Command System. Currently increases efficiency"*.

4.4.7 Future operational use of automated systems

Participants identified increasing capability as the key theme when discussing future operational use of automated systems. Four sub-themes emerged within this theme: sensor capability, capability relating to personnel, tactical decision capability and weapon system capability.

The increasing use of automated systems in future maritime operations will allow the RN to *"have the greatest operational & strategic effect on global influence"*. Across

[1] Warship Electronic Chart Display & Information System
[2] A computerised navigational aid

the data set all participants referred to the critical role automation will continue to play in RN operations; increasing capability in a number of ways. For example increasing the speed at which tactical decisions can be made: *"The speed of reactions required to combat modern threats (esp missiles) is such that an automated system is likely to have a better chance of defeating them in the future"; "in order to speed up processes and/or reduce risk to personnel"; "Command systems and weapon systems require a degree of automation otherwise the speed of response would not be sufficient to defeat threats"; "Clausewitz said "A great part of the information obtained in war is contradictory, a greater part is false, and by far the greatest part is uncertain" - automation will allow vast quantities of information to be presented for the average [human] to understand in a rapid time frame"*; reducing workload of personnel: *"The capabilities of modern weapon systems, and the numbers in which they can be deployed, creates an environment and threat profile that the human mind may be unable to process. We alone, assess, classify, target, engage, as required, repeatedly over a prolonged period."; "The complexity of modern naval warfare can be made easier to exploit by utilising automated systems to remove time consuming tasks away from individuals and allowing more time for thought & decision making"; "In any part of the job as help for the operator in order to reduce the workload by using such tools/systems to execute simple & repetitive tasks"*; increasing sensor capabilities: *"the ability of our sensors to associate what may appear to be background 'noise' to a human operator with a threat"; "Data manipulation and processing in order to populate Command Systems with operational data"; "To allow the operator the thinking time to make decisions on the information rather than process it"*, and greater automation within weapon systems: *"Multiple drones vs a single shipborne helicopter"; "Radars and weapon systems: more effective and potentially make process quicker"; "Automated systems will play an essential role in future warfare when it comes to kinetic action and self defence"*.

Woven into discussions of capability is the second theme of safety. This theme incorporated the safety of the vessel: *"Automated systems for navigation, for example, are of great value to a bridge-watchkeeper- it allows them more time to think, act and surveil"; "In peacetime, for maintaining safety at sea is of paramount importance in my opinion using the example of WECDIS"*; as well as safety with regards to personnel: *"automation is free from crew fatigue, can operate multiple systems from one platform creating cost efficiency and reduce risk to personnel"; "Offboard systems enable commanders to de-risk operations by reducing the hazard to human force elements"; "Areas where it is difficult for a human to work, such as underwater. A replacement for humans that are required for dangerous jobs or where a human cannot work fast enough to complete the job"*. Additionally, SME showed an awareness of the caution that must be applied with the use of automated systems. As one SME wrote, *"Of course there* *is danger in relying heavily on technology"*. Other experts referred to the requirement that automated systems be *"properly tested to ensure that accuracy is maintained and trusted"*, as well as the fundamental need for *"the human [to have] the final veto"* to ensure that current ROE are observed; *"I do not believe that we will remove the 'human factor' in any of our lethal/non-lethal strike options until AI is much more advanced"; "there will still need to be a human decision maker in the chain (particularly the 'kill-chain')"; "There will always be a requirement for human-interface when judgement is required. i.e. we could kill the enemy but should we?"*.

A smaller theme to emerge from the data set were comments relating to system design. SME referred to automated systems *"accuracy is improving all the time"* and showed acknowledgement of the prevalence of *"more sophisticated systems increasing"*. However, it was again highlighted by the SME that caution should be taken with utilising automated systems. As one SME wrote, *"they can simplify highly complex sets of options to provide a smaller selection. They can process a lot of information quickly & present useful results. They can also remove too much information e.g. genuine contacts on radar not displayed due to automated processing"*. Suggesting a level of awareness from SME of the capacity and limitations of current technology. Additionally, for newly developing automated systems to be utilised in the field effectively the reliability of the systems needs to improve, as well as training around what the system is capable of, how it functions and how the operators can supervise the functionalities of the system.

A second smaller theme also emerged from the data with comments relating to financial pressures on the RN. SME are highly aware of the current reduction in manpower, *"The overarching intent in the naval service of automation is to reduce the manpower overhead resulting in shrinking organisational mass"; "the significant manpower issues being experienced, automated tools and systems provide options for lean manning in the future";"[automated systems] can process large quantities of information & cannot be killed, nevermind, are cheaper than the wage bill"; "with the lack of manpower in the RN, I am sure a greater reliability will be on automated services"*, and the impact this has on the requirement to employ automated systems to maintain operational capability.

4.4.8 Exploration of the level of engagement between practitioners and system designers in the development of new automated systems

Of the total number of SME included in this study (N=45), only eight had been consulted in the development of a new system during their career. Table 4.4 presents the eight SME experiences in the consultation process. Highlighting the range of systems SME were consulted on to improve the capability of RN operations. For example, communication systems, mine countermeasures and navigational systems. This reflects the range in experience SME who took part in this study had.

Table 4.4: Comments provided by the eight SME who had experience of being consulted in the development of new systems. [. . .] indicates where text was illegible

SME 2	*"Most systems are not focused around the operator because they are procured by support branches (me/we). When I have been consulted the translation of this feedback into changes have often been selective and greater reflect the opinion of those procuring them. When I have had the opportunity to procure kit myself to support operations the impact is transient as the tactical development achieved cannot easily be fed back into the organisation. The organisation has considerable change inertia issues."*
SME 8	*"Yes. I was a very small contributor to the ECPINS/NAV system for HMS Queen Elizabeth Class in its very early stages of development. Given my currency & previous experience as a navigator"*
SME 13	*"I tested the comms for the QE "*
SME 21	*"Development of MCM Expert"*
SME 31	*"Yes, but not often- as an operator usually hardware is pre-purchased. Occasionally I have had input into software developments based on user feedback"*
SME 35	*"Yes, NAUTIS replacement"*
SME 40	*"Once but it received no feedback. It was for the new MCMU [. . .] Indicator"*
SME 41	*"Yes. I was consulted by EIPT from BAE with regards to the new UK/FR MCM System. Whilst serving as the XO in HMS Cattistock. I was also consulted on the installation of the Remms 600 RUV platform onto Hunt MCMVs."*
QE	HMS Queen Elizabeth
MCM	Mine Countermeasures
MCMU	Mine Countermeasures Unmanned
EIPT	Equipment Integrated Project Team
XO	Executive Officer
MCMVs	Mine Countermeasures Vessels

4.4.9 Exploration of the opinions of RN personnel of the current consultation process

Three themes emerged through exploring SME views on the consultation of current and future personnel during the design and development stages of new automated systems (consisting of 49 nodes and approximately 1,189 words); Collaboration (36.7% of total nodes), Knowledge transfer (28.6%) and Fit for purpose (34.7%).

Emerging from the data is a clear desire from current personnel to be consulted in the design and development stages of new automated systems. As one SME wrote *"How can anyone design an automated system without consulting the people who will use it?"*. The concept of collaboration between personnel and software engineers appears to stem from an awareness from SME of the requirement that *"to develop kit you need to use operator knowledge as well as technical officers to develop new system"*. The collaboration and the transfer of knowledge that results from this communication ensures that systems are built fit for purpose: *"It is very important that users are consulted in the design process in order to ensure that the final product is user friendly and fit for purpose"*.

However, the data also highlights that this collaboration seldom occurs: *"Consultation seems to take place engineer to engineer & not include the operator/end user"; "The end user is rarely consulted for requirements & during development pieces"; "Terrible. Only those within teams are consulted and they become an echo chamber of ideas. The only people who should be asked for input at the design stage and in the testing phase are the operators"*. This lack of collaboration results in systems that are not fit for purpose, *"many current tools are not fit for purpose as there was no operator input in the design"; "Kit designed and built by engineers having never been employed in a maritime environment always has problems"*. A key function identified from SME and the wider literature of introducing more automation into military operational environments is to improve capability. However, if systems are not being designed in collaboration with the personnel who will be using them daily, this results in the likelihood that new systems won't be used to their full capacity: *"It is essential that those who are going to actually use these tools or systems are consulted as then the system/tool will actually be usable, rather than being overly complicated"*.

This disconnect impacts upon the capability new systems can provide RN personnel and can have temporal and financial implications: *"Too often once we get to use new equipment it quickly becomes clear that one or two main details that if better thought out could have made an enormous difference. Such is the nature of equipment programmes that it is incredibly hard/slow to change such things once in service"; "Generally poor integration of change programmes into operator domain. Contractor buoyancy and drive often overrides operator concerns. Once accepted into service these issues endure. Often the timeframe for installation does not support rigorous change process especially when a limited number of platforms prevent full tactical development and testing prior to operational employment"*. Part of this disconnect may be due to lean manning of current RN operations. As the previous quote highlights, if designers are unable to test their systems in ecologically valid environments, for example on the platforms they will eventually be used on, it is easy to see how tools may not be fit for purpose when first released.

Although challenges exist for system designers to test their product in ecologically valid environments this should not prevent the engagement of current personnel in the planning and initial concept design phases: *"You must consult the operator of equipment (at varying levels) to determine gaps in technology that need to be improved and then to put possible amendments of fixes to the system forward for expert, and current, opinions on those improvements"*. It is common for contractors to consult their own in-house experts when designing new systems however as one SME in this study pointed out, *"Whilst people may have had previous experience to contribute, it may have been some time since they had been at sea as an "end-user" with other advances in technology or a full understanding of the needs/requirements"*.

Collaboration is crucial to developing new systems that are fit for purpose. The knowledge transfer that must occur to enable this is not just from operator to engineer but it is also important for operators to learn the capacity that current and emerging technology has: *"current serving personnel have very little experience beyond their current equipment and rarely have much knowledge of current/emerging technology. Thus the current community is often not well placed to advise on new automated systems and tools when they have little exposure or experience of them"*. An awareness of these limits to their knowledge was shown by several SME, eight of whom omitted to answer this question: for example, *"This is not my professional field; currently I am of the operator level of tools/systems"*. One way in which operators may be more willing to comment on what features of systems would be useful to their job roles is through training: *"Training will be needed"*. Alternatively, if collaboration becomes commonplace within the organisation, the transfer of knowledge this will allow will result in both operators and engineers developing a common language (Robertson et al., 2003). This will facilitate understanding of each others domains of expertise.

4.5 Discussion

Data was collected from N=45 experienced RN personnel to explore the current use of automated systems, where automated systems may be used in the future and the consultation between system designers and current personnel. The data also revealed a desire from SME to be involved in the design and development stages of new systems. However, there exists a lack of collaboration and knowledge exchange which currently results in systems not always being built fit for purpose.

SME identified the use of automated systems across all aspects of operations onboard RN vessels. This is in line with current literature that posits the increased occurrence of socio-technical systems within military environments, for example see Ministry of Defence (2011) and Ministry of Defence (2014). SME also provided a comprehensive picture of where automated systems may play increasing roles in future operations. The nature of warfare will change as automated systems and the use of unmanned vehicles increases. It has been argued that as humans may have less front-line involvement with combat this may change public and political opinion towards combat, which could increase the likelihood it occurs (Ministry of Defence, 2014). It is therefore of paramount importance that all automated systems and unmanned systems are developed fit for purpose. The data presented shows that this is not currently always the case.

Automated systems are required to perform alongside individuals and teams in complex environments. Field research has highlighted how individuals and teams self-organise and continuously adapt to their current situation - at times diverting from standard operating procedures (Bigley and Roberts, 2001). Therefore, as highlighted by Naikar (2018), systems designed according to normative procedures may hinder the ability for sociotechnical teams to self-organise when faced with complex, dynamic and time constrained environments. The inability for systems to adapt increases the operators cognitive burden as they will have to adapt not only to the situation but also to the constraints of the system. This requirement was highlighted by experts in this study. Therefore, it could be that personnel viewed systems as not fit for purpose as the systems were unable to adapt to the challenging environments the sociotechnical team faced.

The percentages of quotes covering current automation use (25.6%) and future use (54.3%) suggest that current personnel have a wealth of ideas that could be tapped into by system developers. Personnel also showed great awareness of the capabilities of technology, where systems will have the most benefit to their day to day operations but also the caution that must be taken when using these systems. However, the data highlighted the limited number of current personnel (n=8) who had been previously consulted in the development of new systems. Where consultation has occurred, it is questionable how the input from practitioners was received and if it was taken on board by the system designers. RN personnel are of course not experts in software development, the SME showed an awareness of this, however, they are on the front line using the systems that are developed. The data also showed that when systems are brought in that they do not function as expected or are not trusted by the personnel or their commanders, this results in increased workload for RN personnel. To facilitate improved future collaboration between personnel, industry and academia adopting an approach like the Athena project may provide an open environment where ideas and knowledge can be exchanged.

Increasing the reliability of new automated systems has been a key feature of human factors and ergonomics research. A wealth of literature highlights the importance of operators' awareness of the reliability of systems to ensure appropriate use and facilitate trust in the system (Hoff and Bashir, 2015; Madhavan et al., 2006; Parasuraman et al., 2012). This body of literature has also shown how trust in automation is influenced by a complex combination of factors, for example individual differences in working memory (Parasuraman et al., 2012). Therefore, trust is not formed solely based on the features of the system (Hoff and Bashir, 2015). Research has also shown how reduction in trust due to reliability errors can be mitigated if the operators are provided with information as to why the system may err (Dzindolet et al., 2003). RN personnel could gain this information through being involved in the development process of new systems via communicating with system developers.

The data further showed how current personnel voiced caution towards over-relying on automation. Recent collisions involving US naval vessels and merchant ships have highlighted the dangers of over-relying on automated navigation systems (Forcast International, 2017; Fraher, 2017). A complete overhaul of traditional navigation procedures in favour for automated systems has led to crews being insufficiently skilled in basic navigational procedures. Which has led to a reduction in their vigilance of the surroundings, relying instead on the automated system to navigate their path. Caution voiced by RN personnel in this study is well founded. Additionally, with the UK's current ROE a human will always remain *in* the decision-making loop. Therefore, personnel are continually adapting to their new roles as fully trained and capable naval officers as well as supervisors to increasingly sophisticated systems.

SME also highlighted the limited number of platforms that are available and the limited routes by which new systems can be tested prior to their release into service. One way in which academia could provide support for this problem is through the continued development of immersive and ecologically valid test-beds that generate domain specific challenges (Jenvald and Morin, 2004). The value of training personnel within immersive environments has been shown in several fields, from the military (Jean, 2008), to emergency services (Alison and Crego, 2008), to medical practice (Kirkman et al., 2014). Further, by providing personnel with safe-to-fail environments (Rouse, 1991), cognitive learning can be facilitated (Klein and Baxter, 2006). The application of immersive environments for testing new automated systems can provide a way in which systems can be robustly tested to ensure they translate across into the operational environments. Additionally, the dual-purpose approach to training personnel and testing new systems could be enabled by combining these two goals into one immersive environment. The data generated by SME who took part in this study suggests that an innovative approach to system procurement is needed to ensure that capability is maintained across RN operations.

Combining training of personnel with testing of new systems would provide an environment that fosters collaboration and knowledge exchange between practitioner and designer. There are of course challenges associated with the development of immersive realistic test-beds for military operations, for example security concerns. To address these challenges current industrial partners, hold the capabilities to test new systems in highly secure immersive environments with high-fidelity scenarios. However, such test-beds are currently not easily accessible to academia or other industries who may be researching and developing new systems. Additionally, there is little ability to quickly bolt on new systems or amendments to current systems to perform quick tests of their functionality and how operators interact with them. Personnel are also restricted on their availability to take part in test research due to the reduced manpower the RN currently operates with. Therefore, building stronger links to academic partners who currently have test-bed capabilities, for example The Command Teamwork Experimental Test-bed (ComTET). [3] could provide an alternative route to system development within academia. Immersive simulations can elicit how systems are interacted with during operations, showing their flaws and strengths. Personnel will also develop an understanding on how the systems function, their capacity and reliability through taking part in training. This understanding can facilitate more accurate system use and mitigate against misuse, disuse or abuse (Parasuraman and Riley, 1997). An additional benefit to academic links is the ability for university students to be initial participants in studies, providing larger sample sizes to test initial developments on. These initial development studies will allow features to be statistically explored prior to running experiments with expert personnel.

Limitations

A potential weakness of using a questionnaire based method is that it was not possible to probe upon answers provided by participants. However, the findings from this study can be used to formulate topics to cover if researchers have greater access to SME to conduct more comprehensive interview techniques, such as the Critical Decision Method (CDM). It is also acknowledged the shortcomings of qualitative research to provide generalisable conclusions. The uniqueness and size of the sample arguably

[3]ComTET was designed to explore how information flows within socio-technical command and control teams (Roberts et al., 2015)

mitigates this limitation. The (N45) RN personnel who completed the questionnaire encapsulate a range of experience covering both underwater and above water job roles. Although organisational cultural norms may arguably influence the responses provided, it is beyond the scope of this paper to explore this. However, it would be of great interest for future work to explore the influence cultural frames can have upon responses (Klein, 2004). Additionally, research that draws upon eliciting knowledge from experts is often considered vulnerable to bias in interpretation of the data (McAndrew and Gore, 2013). However, the use of inductive data analysis techniques such as Thematic analysis, applied using a transparent framework can reduce this vulnerability.

Implications for findings

This chapter presents unique insight into currently serving RN personnel's views of the existing and future state of automated systems and the present situation surrounding the procurement of new systems. The privileged access to this sample population is a rarity and combined with the range of experience provided by each SME and the size of the sample (N=45) provides a singularly exclusive window into a hotly discussed research topic. Alison et al. (2015) argued for the importance of credibility and transferability of conclusions gained from naturalistic research as a way to judge research. The quotes provided by experts woven into the analysis presented, provide strong grounds for highlighting the credibility of this research. Although looking specifically at RN operations, the recommendations provided by SME arguably provide transferable conclusions. Globally the decisions made within naval operations are similar, with additional factors of fleet capacity, manpower and available technology influencing standard operating procedures. Research into human-machine-interaction has highlighted the facilitators and barriers to appropriate automation usage; these often transcend operational environments and cultures. Therefore, the results presented in this chapter can provide avenues for future research to explore with naval organisations globally.

Conclusions

Currently a disconnect exists between RN personnel and system designers. This results in new systems not always being fit for purpose which can negatively impact upon RN operations across all domains. For example, by increasing the workload of personnel who are already balancing time pressures, uncertain information and the criticality of the environment. To overcome this, greater communication and knowledge exchange should be promoted between currently serving RN personnel and system designers. One way in which this could be achieved is through open forums where personnel could suggest new systems that would support and improve their operations. For example, a forum similar to the Athena project. Additionally, to ensure personnel have an understanding of current and emerging technology, system designers could also present their ideas at this open forum. A collaborative environment such as this would enable knowledge exchange and promote the development of systems that are fit for purpose at time of release. It is worth remembering that *"war is ultimately a human endeavour. It will be humans who choose to go to war, it will be humans who can stop wars and it will be humans who suffer the consequences of war"* (Ministry of Defence, 2014, p. 96). Therefore, the humans on the front line should be consulted in the way developments are made.

The next step of this thesis is to present the development and testing of the microworld Automatic Radar Classification Simulation (ARCS). Accordingly, the next chapter discusses the pilot tests conducted to check the usability and validity of ARCS followed by Chapter 6 which presents the mixed-design student experiment.

4.5.1 Acknowledgements

The quotes provided to support the analysis are the views and opinions of the personnel who completed the questionnaire. They do not necessarily reflect the official policy or position of the Royal Navy or Ministry of Defence.

Chapter 5

The design and development of ARCS

5.1 Abstract

This chapter describes how the Automatic Radar Classification Simulation (ARCS) was developed and tested with two small pilot studies. Pilot 1 (N=6) and Pilot 2 (N=5) were conducted at the University of Liverpool with PhD students taking part as participants. ARCS has been designed with several adaptable features, for example the ability to edit and create scenarios, edit the functionality of the automated system, and incorporate additional questions or surveys into the task. The pilot tests showed that the required performance metrics; total number of correct and incorrect classification decisions, selection of system, and workload ratings were all collected by ARCS. Additionally, the targeted workload dimensions (mental and temporal demand) were elicited by the scenario. However, several recommendations were produced through these tests, for example, to provide participants with a longer and more detailed training brief. These recommendations were then implemented prior to the main experiment. Overall, ARCS was found to be a fully functional microworld that can be edited and adapted to meet the requirements of the researcher.

5.2 Introduction

The previous chapter illuminated the ubiquitous use of automated systems across all operational aspects of the Royal Navy (RN), for instance from navigation to the identification of hostile threats. Automated systems are proposed to support decision making and ensure that RN personnel complete tasks effectively. However, there is a clear disconnect between those designing and building systems with the personnel on the front line who will be required to use them. Additionally, clear concern has been identified with the regularity of new systems being brought into service that are not fit for purpose. One way in which this disconnect can be bridged is through closer collaboration between RN personnel and system designers. This human-human collaboration, when designing and developing automated systems can be supported by academia. Academics are well placed to develop open architecture high-fidelity testbeds that new systems can be robustly tested with prior to their release. Utilising immersive environments will allow RN personnel to be included in the testing phases ensuring the systems are fit for purpose at time of release and throughout their lifecycle to ensure continued operational support to front line personnel. Accordingly, the next section of this thesis describes the development of a novel microworld test-bed that was designed by the author of this thesis and developed in collaboration with a software engineer to explore automation usage decisions in a naturalistic way. This chapter aims to:

(i) Provide a brief review of the testbed development and experimental approach

(ii) Describe the adaptable features of ARCS

(iii) Present the findings from the pilot studies and the changes made prior to the main experiment

Researchers continue to face the challenge of maintaining experimental control when exploring complex real-world problems. A constant trade-off between ecological validity and experimental control has to be made when designing a research approach (Loomis et al., 1999). When researching complex environments, such as military command and control decisions, there is increased difficulty in maintaining experimental control (Dörner and Funke, 2017). One way in which to manage this trade-off is with the use of virtual simulated computer environments. Employing naturalistic research methods such as the critical incident technique (outlined in Chapter 2) allows the researcher to develop complex scenarios that can then be utilised within simulated computer environments (Reuschenbach, 2008, cited in Dörner and Funke (2017)). One example of the application of a virtual simulated computer environment is in exploring submarine command team functionality (Stanton and Roberts, 2017).

In line with this, ARCS was developed by the researcher to provide a simulated computer environment that mimics the operational environment of the development of the recognised air picture. Compilation of an accurate recognised air picture is crucial to enabling the predication of engagement outcomes as valid (Foard, 2000). As previously discussed in Chapter 2, the process of developing ARCS involved 4 steps; (i) building knowledge of the specifics of the research domain, (ii) designing prototype interface and developing task structure, (iii) piloting the system, and (iv) recruiting participants and utilising ARCS. This chapter addresses step three, piloting the system. Pilot tests are a common methodology employed to conduct manipulation checks on a new experimental design or testbed (van Teijlingen and Hundley, 2001). As ARCS was developed specifically for the research presented in this thesis, two pilot tests were conducted to determine the validity and utility of ARCS as a microworld, a brief overview of each pilot test will now be discussed.

5.3 ARCS Development

5.3.1 Materials

ARCS was developed by the author with collaboration from a computer programmer as a Java run high fidelity testbed. The open architecture of ARCS enables the test-bed to be adapted to test a variety of research questions. The adaptable features of ARCS will each be discussed.

Scenarios

The scenarios presented during the training and main task phases can be edited and changed. For example, this study explores development of the recognised air picture in a littoral environment. However, the geographic map depicted within the main task is a fully configurable .jpeg file and therefore can be changed to depict any location desired by the researcher. The speed at which the scenario runs can be increased or decreased depending on the research aims. For this study the scenario was run in real time to promote the realism of the task. However, task difficultly could be raised by increasing the speed at which the scenario runs.

To facilitate psychological immersion in the task it is possible to play an audio track alongside each scenario such as air traffic control chatter. For this study, the audio track did not relate to the track movements on screen, this was due to time constraints with setting up the experiments. However, future research should look to utilise the audio feature to match the scenarios further increasing psychological immersion.

Currently ARCS runs as a standalone test-bed, the scenario generated is designed to explore how an individual would perform when classifying all radar tracks on screen. However, within the RN the recognised air picture is developed by a team of operators. Therefore, although not working together, participants of the study presented were run in groups to simulate the working environment in a Type 45 Destroyer operations room. Future versions of ARCS could be developed to perform specific aspects of this classification task, enabling teams of participants to complete the task together.

Track generation

The tracks incorporated into ARCS can be loaded through ADS-B collected data on location, altitude and speed or the researcher can manually generate tracks. This feature makes ARCS highly adaptable and allows the creation of real-world track profiles.

As ARCS has been developed using the ADS-B standard it allows the micro-world to potentially be deployed in a live operational environment. For the study included in this thesis, a combination of the two was used. All neutral tracks, i.e. commercial aircraft, were real flight patterns collected from freely available ADS-B data from websites such as flight tracker [1]. All friendly and hostile tracks were generated by the researcher who had gained an understanding of the flight profiles associated with friendly or hostile tracks from undertaking the research presented in Chapter 3 and through informal conversations with experts.

Automated system

The automatic classification system can also be edited in several ways. The number of times a participant is able to select the system can be set at the start of the experiment. The length of time it takes to classify a single track and how long it will run for when selected can also be set. The automated system used in this study only classified neutral tracks and worked alphabetically down the list of tracks included in the scenario. However, ARCS is capable of handling bolted on systems in order to test sophisticated classification algorithms.

Additional measures

Integrated into ARCS is the ability to pause the current scenario to present participants with questions. This study utilised the NASA TLX (Hart and Staveland, 1988) to provide accurate workload ratings throughout the experiment. The NASA TLX is a commonly used assessment tool to collect workload ratings across six dimensions: Mental demand, Temporal demand, Physical demand, Frustration, Performance and Effort. Defined as *"a term that represents the cost of accomplishing mission requirements for the human operator"* (Hart, 2006, p. 904), understanding operator workload is a crucial factor when designing automated systems.

The open architecture of ARCS enables any form of ratings or questions to be incorporated into the main task depending on the research aims. The timing of when these ratings will be shown and how often can also be edited.

5.3.2 Data collection

ARCS records all mouse clicks made by participants. This enables a wealth of quantitative data to be collected. For example, it is possible to collect:

- All classification decisions made by participants and the accuracy of those decisions
- The time between a track appearing on screen and it being classified

[1] https://www.flightstats.com/v2/flight-tracker/search

- The location of each track when it is classified
- The number of times the automated system is selected and how many classifications it makes
- A second by second tracking of all track movements on screen
- A second by second analysis of decisions made by participants. For example, what strategies were used to approach classifying all tracks on the screen, did participants start on tracks located closer to their vessel and work away from their ship?

5.3.3 Procedure

The experiment consisted of 3 stages: (1) the Training Session, (2) the Main Task, and (3) the Focus Group. Each stage will now be discussed in more detail.

(1) The Training Session

Following obtaining informed consent, participants were taken through the 25-minute training session which consisted of a pencil and paper exercise followed by a familiarisation task. Participants were shown the iconography used within ARCS and provided with a worksheet to fill in the meaning of each symbol (see Figures 5.1 and 5.2). This worksheet could be referred to by participants throughout the experiment to ensure comprehension of each symbol. Participants were also provided with information to aid the classification task, for example, a commercial fixed wing aircraft could be assumed to travel at an altitude between 20,000-40,000 feet and at a speed of up to 500 knots; unknown to participants all commercial tracks were neutral (see Table 5.1). Finally, participants were provided with definitions for the NASA TLX workload ratings, as is recommended practice for using this workload measure. All information sheets provided to participants can be found in Appendix Four.

Table 5.1: Information provided to participants to aid their classification decisions

Platform	Altitude	Speed
Commercial Fixed Wing	20,000-40,000 ft	Up to 500 knots
Commercial Rotary Wing	Up to 8,000 ft	100-200 knots
Military Fixed Wing	14,000-25,000 ft	Up to 800 knots
Military Rotary Wing	Up to 10,000 ft	100-135 knots
If information relating to a track meets these criteria you can assume the relevant platform		

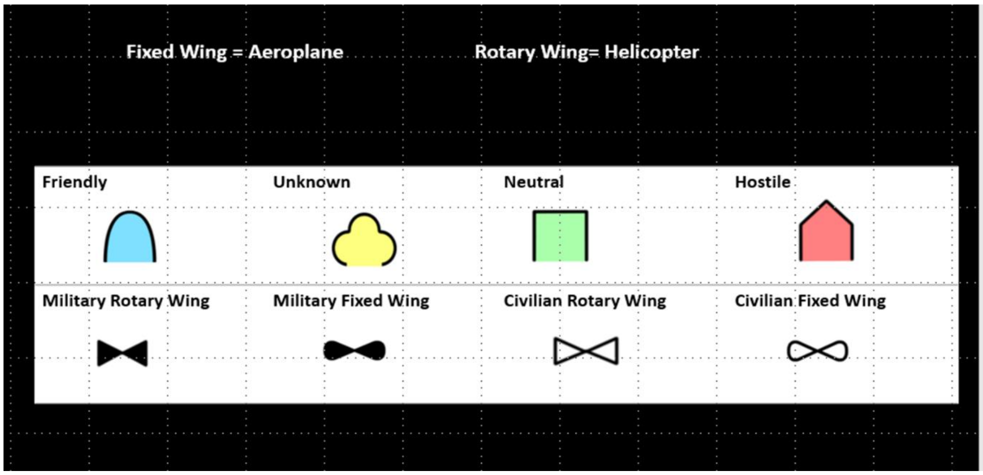


Figure 5.1: Slide taken from the training presentation depicting the symbology used in the simulation (The symbology was taken from NATO Standarsization Agency (2011))

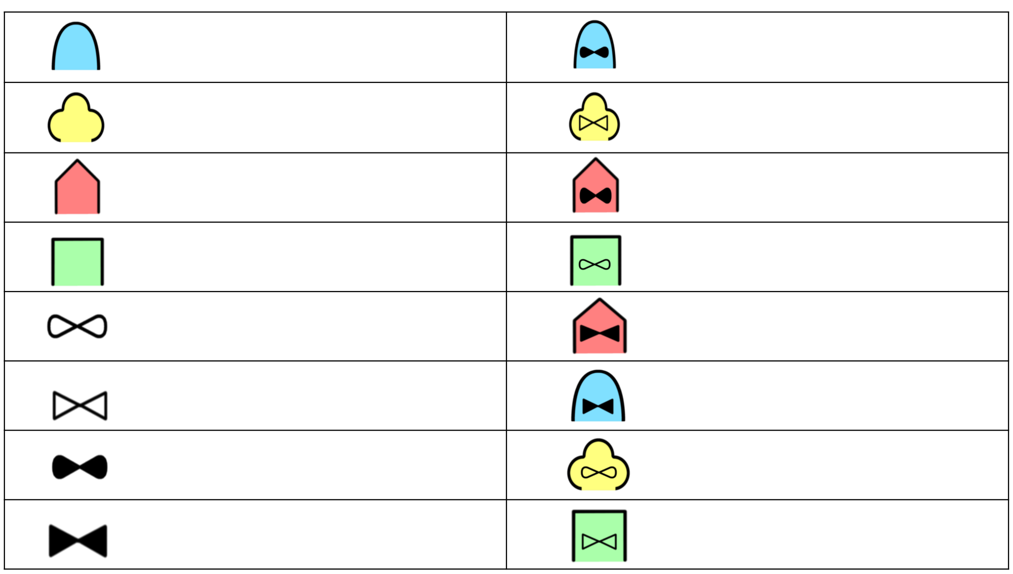


Figure 5.2: Section of the training sheet participants completed

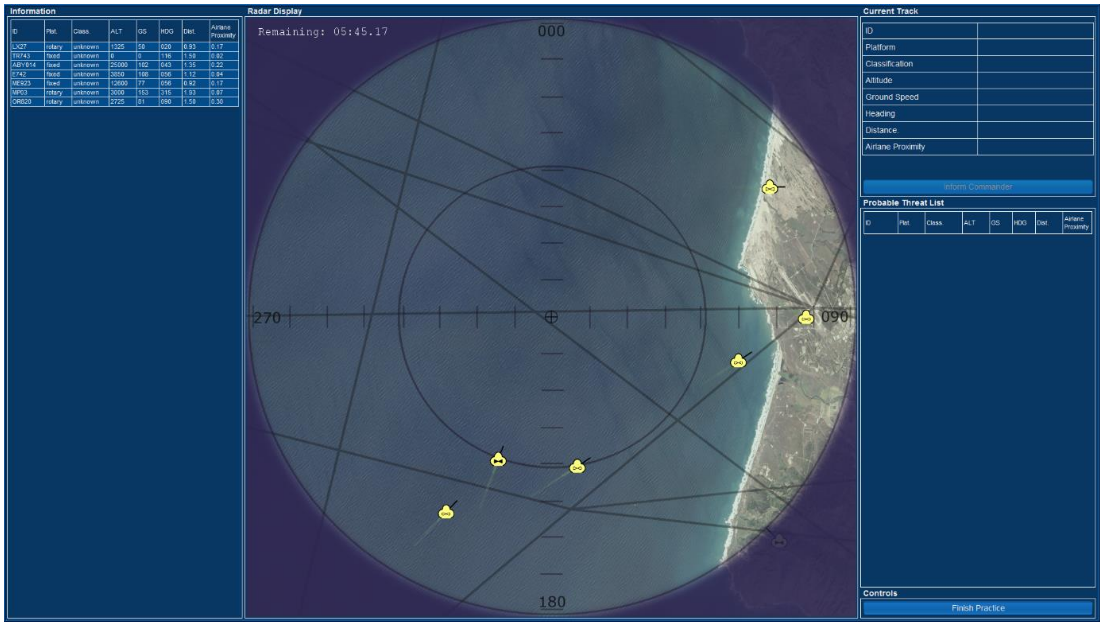


Figure 5.3: Screenshot of the training session with ARCS

Next participants completed a 15-minute familiarisation task with ARCS (see Figure 5.3). During these 15 minutes participants were asked to classify all tracks (icons) on the radar picture into friendly, neutral or hostile. Open source information (such as ADS-B data- including altitude, velocity, heading, proximity to airlane etc.) relating to each track was presented to the participants to aid their decision-making process. The training section took 15 minutes to complete, with participants experiencing a range of difficulty levels (e.g. a range in number of radar tracks) to ensure that participants became comfortable with the task itself and the testbed interface.

(2) The Main Task

Following the training session participants were given the opportunity to ask the researcher any questions before re-viewing the 'Daily Tasking Orders' in front of them (a paper based inject, highlighting their own responsibility (high accountability) or diffusing the responsibility (low accountability) for any incorrect classification decisions). Participants then completed a single scenario that ran for 30 minutes (see Figure 5.4). During which time participants were required to classify all tracks on the radar screen into friendly, neutral or hostile. If any tracks were deemed as potentially hostile, the participants were required to click the 'Inform Command' button to flag those specific tracks. At any point in time the participants could also opt to employ the automated classification tool. For Pilot 1, passing control to the tool could be handed over and taken back by the participant wherever and as many times as they saw appropriate. For Pilot 2 participants could only select the automated system twice and the system would run for 10 seconds before giving control back to the participant. This change in system operation was based on feedback collected at Pilot 1. At 5-minute intervals the scenario was frozen and the NASA TLX ratings were presented to participants (see Figure 5.5). Following the completion of the ratings the simulation would unfreeze and participants continued with the main classification task.

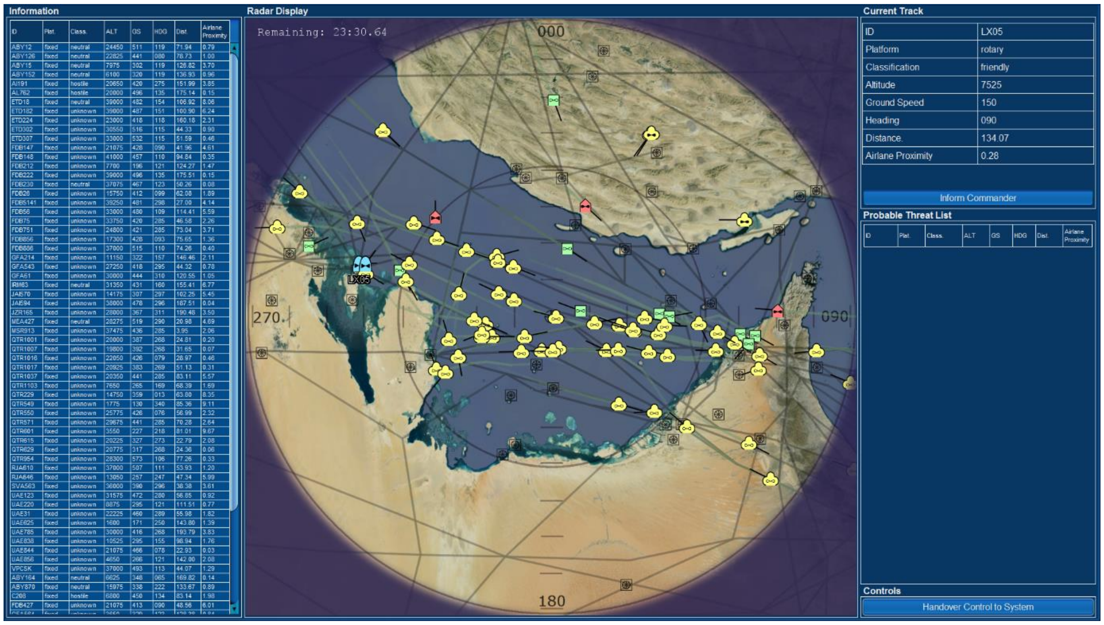


Figure 5.4: Screenshot of Pilot 2 the main task within ARCS

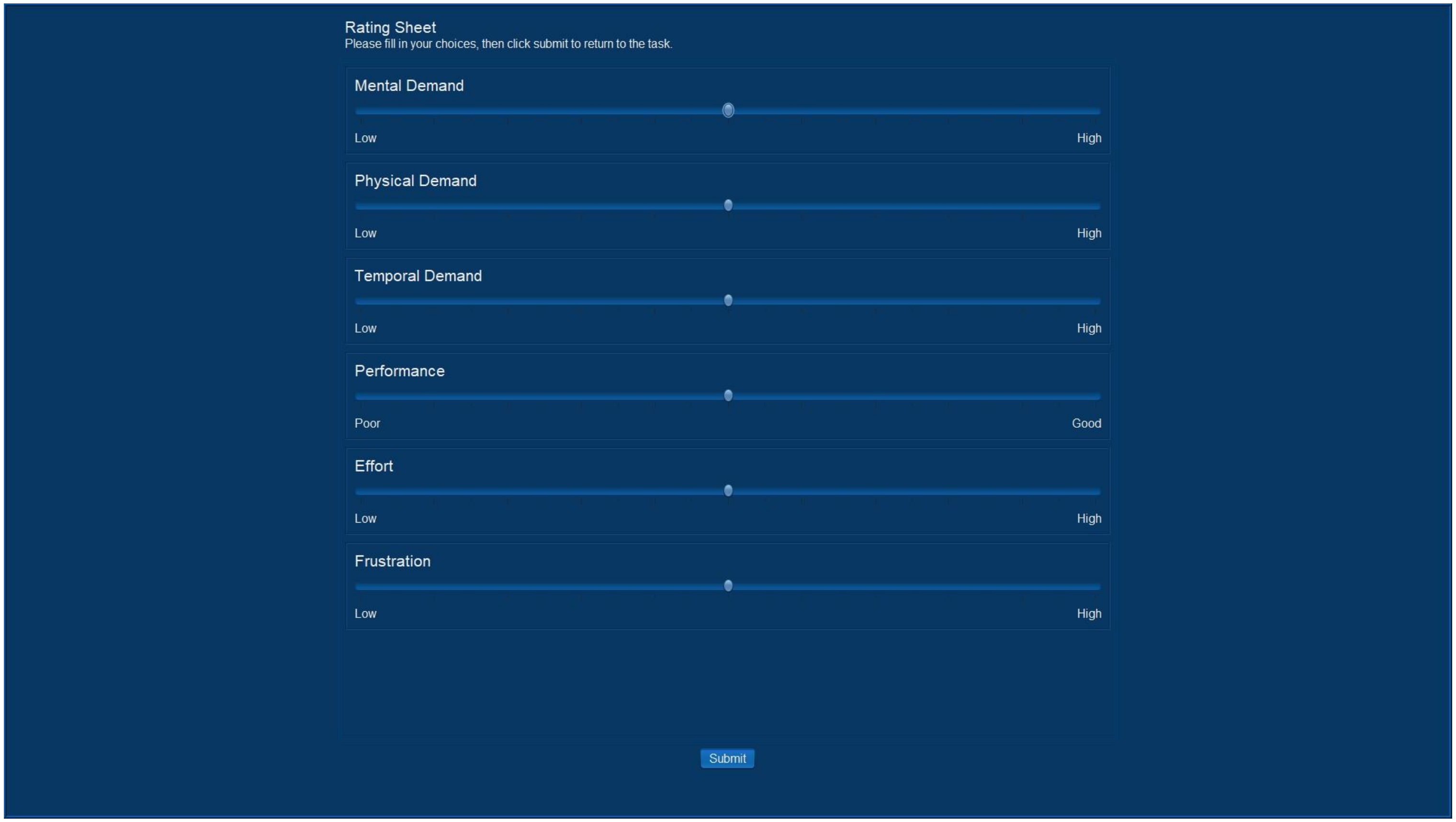


Figure 5.5: Screenshot of NASA-TLX ratings participants completed at 5-minute intervals throughout the main scenario

(3) Focus Group

Having completed the main task participants were asked to complete six questions (see Table 5.2) to provide feedback on perceived performance and rational behind decisions made. Specifically, the rationale behind when and why participants employed the automated support tool. Participants were also asked to complete six cognitive trait measures: Cognitive flexibility inventory, Need for closure, Assessment & Locomotion, BIS-11, Conscientiousness, and Propensity to trust (Chapter 7 discusses the selection of these measures in more detail). Finally, a focus group was held to delve into the participants' feedback. Each stage of the experiment was discussed giving participants the opportunity to raise any issues or positive feedback regarding the experimental design and the task itself.

Table 5.2: Questions participants completed following the main task

1	From 1-10 how would you rate your performance? (1=poor, 10=perfect)
2	On a scale of 1-10 how accountable did you feel towards the decisions you made? (1=not accountable, 10= very accountable)
3	Did you choose to use the decision support system at any point(s) during the scenario? YES/NO
4	Reflecting on your performance, where there any occasions that you think you should have used the decision support system but did not and why?
5	Reflecting on the scenario, where there any occasions that you felt you could have use the decision support system earlier but did not?
6	What made you not choose the system at an earlier point?
7	If you could perform the scenario again, is there anything you would do differently?

Prior to this pilot test, ARCS was shown to the Military Advisor (MA) and Subject Matter Experts (SME) at Dstl. The demonstration consisted of the researcher going through ARCS in detail with a SME. Providing comments and feedback on the fidelity, both ecologically and psychologically, of the system interface and the task participants would be required to complete. Detailed notes of this meeting were taken by the researcher and will be woven into the results and discussion.

5.3.4 Pilot Study Participants

N=8 participants took part in the pilot studies (see Table 5.3 for breakdown of participant characteristics). n=3 participants took part in both pilot studies with 11 weeks between each pilot test. This allowed the researcher to explore the effect the changes made to ARCS following Pilot 1 had upon participant experience and performance. Having both experienced and novel participants in Pilot 2 enabled an understanding of how both groups of participants interacted with ARCS and their overall performance at the task. Ideally, more participants would have been recruited for these two pilot tests, however, due to time constraints small sample size numbers were accepted with an understanding of the limitations upon data analysis.

Table 5.3: Breakdown of participant characteristics across both pilot studies

Pilot 1	Pilot 2
N = 6 3 = Male, 3 = Female	N = 5 2 = Male, 3 = Female

5.4 Analysis

As already mentioned the sample size of both pilot tests does not allow an in-depth analysis on task performance. However, the purpose of these pilot tests was to prove the utility of ARCS as a research microworld. Therefore, the results section of this chapter will focus on what data is collected through using ARCS.

5.5 Results

5.5.1 Use of tool

Five of the six participants in Pilot 1 used the support tool to check or confirm their classification decisions (see Table 5.4). Arguably the collective appraisal errors made by participants were due to a lack of understanding relating to the system itself, i.e. that the tool would not check classifications the participants made. Participants were not informed about how the system worked or given examples of when to use it, thereby forcing participants to assume the system function.

Of the 5 participants in Pilot 2, 4 utilised the tool and 1 did not. Pilot 2 comments from participants related to the number of times the system could be used, *"I wish I had more than 2 Decision Support System (DSS) tokens. As once one was used (to test its ability) with only one remaining it was difficult to decide if the difficulty was worth enduring, thinking there may be a more opportune moment in the future"*. Across both pilot tests participants stated that the first use of the automated system was to test its capability.

Table 5.4: Quotes from participants when asked why they chose to use the decision support tool

"I felt unsure about what I was doing"
"Concern that all tracks were classified as neutral, wanted "support" from system to confirm"
"To see how the decision support would decide in order to understand decision criteria and validate choices"
"To see if/how it helped and to see if it amended any of my decisions"
"I wasn't 100% sure if I was adequately classifying so wanted to make sure I was on the right lines. It also provided an opportunity to look at the bigger picture more easily"

5.5.2 Workload

Analysis of the NASA-TLX ratings taken at 6 time points within the main task revealed that participants perceived workload remained relatively stable throughout the task (see Figure 5.6). Mirroring the observed plateauing of actions 10 minutes into the 30 minute scenario. Analysis of the weightings given to each workload aspect globally, revealed that mental and temporal demands contribute more to overall workload, suggesting that the task did elicit the desired cognitive load (see Figure 5.7). However, as the workload ratings did not fluctuate, alterations to the scenario will be made in order to elicit a greater level of workload during the critical time point. Table 5.5 displays the global workload rating for each participant across the 30-minute scenario. Data was lost for 3 participants due to technical problems with ARCS. However, this was fixed before Pilot 2 to ensure that all workload ratings were collected.

NASA-TLX ratings collected during Pilot test 2 show more varied workload throughout the 30-minute task (see Figure 5.8). The global weighting of temporal demand was lower compared to Pilot 1, however mental demand was still rated as the highest contributing factor to workload (see Figure 5.9). Further alternations will be made to increase the workload during the critical event of the main task.

Due to the limited number of participants it was not possible to run full statistical analysis on the influence workload may have had upon tool use. However, it was possible to look at the descriptives of workload and tool use across the six phases of the experiment (see Table 5.6). The workload scores were collected at the end of each phase. Therefore, it is possible to view the change in workload potentially attributable to the use of the automated system. Participant 2 for example, reported high workload during phase 3 when they had also selected aid from the system, their workload increased again in phase 4 and reduced in phase 5 following utilising the tool for a second time. Although cause and effect cannot be ascertained from this data alone, with more data collected a trend between workload and tool use may be elicited.

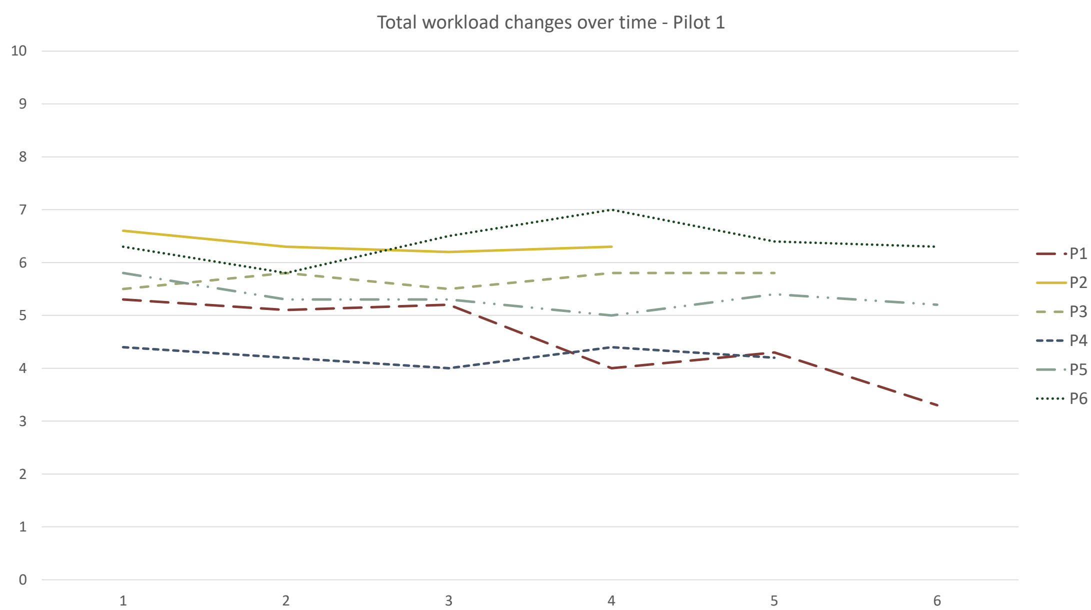


Figure 5.6: Global workload ratings from participants in Pilot 1. Missing data resulted due to technical problems with the recording of data during Pilot 1

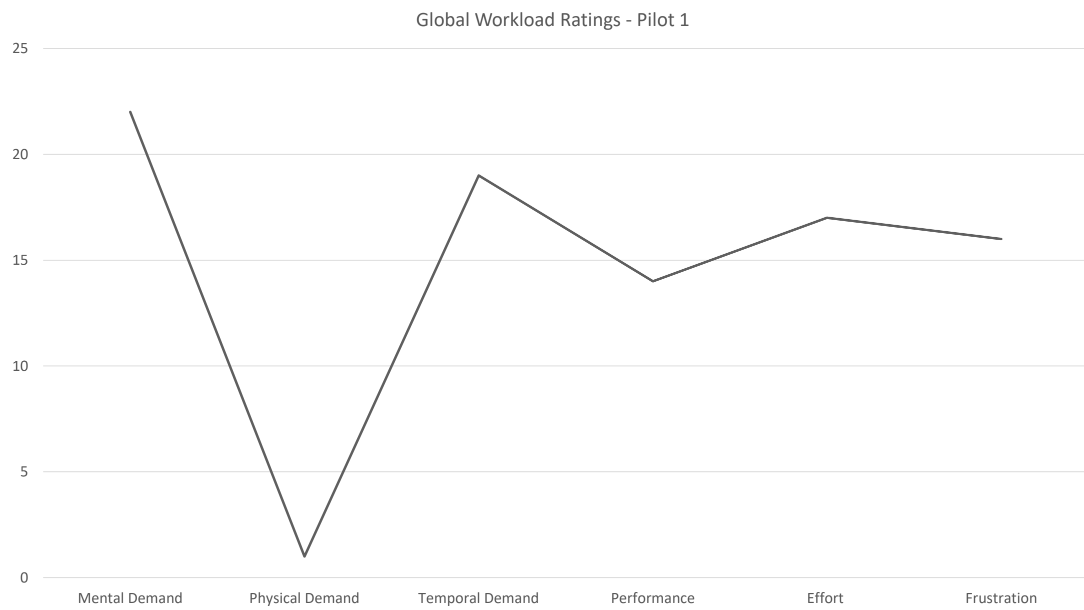


Figure 5.7: The weighting given to each workload measure by participants in Pilot 1

Table 5.5: Descriptives of self-perceived workload during Pilot 1 across the experiment as measured by NASA-TLX

ID No.	Phase 1 (0-5mins)	Phase 2 (5-10mins)	Phase 3 (10-15mins)	Phase 4 (15-20mins)	Phase 5 (20-25mins)	Phase 6 (25-30mins)
1	5.3*	5.1	5.2	4	4.3	3.3
2	6.6	6.3	6.2	6.3*		
3	5.5*	5.8*	5.5	5.8*	5.8	
4	4.4	4.2	4	4.4*	4.2*	
5	5.8*	5.3*	5.3*	5*	5.4*	5.2*
6	6.3	5.8	6.5	7	6.4	6.3

* = used the automated decision support system

Table 5.6: Descriptives of self-perceived workload during Pilot 2 across the experiment as measured by NASA-TLX

ID No.	Phase 1 (0-5mins)	Phase 2 (5-10mins)	Phase 3 (10-15mins)	Phase 4 (15-20mins)	Phase 5 (20-25mins)	Phase 6 (25-30mins)
1	7.9	7.5	7.8	5.9*	8*	5.6
2	7.4	7.8	6.7	7.5	5.7*	7.4
3**	6.3*	6.4	6.1	5.9	5.7	6.2
4**	7.4	6.4	5.8	5.5	4.4	3.8
5**	5.3	5.1	6*	7.4	5.8*	6.6

* = used the automated decision support system
** = had prior experience with the task and system

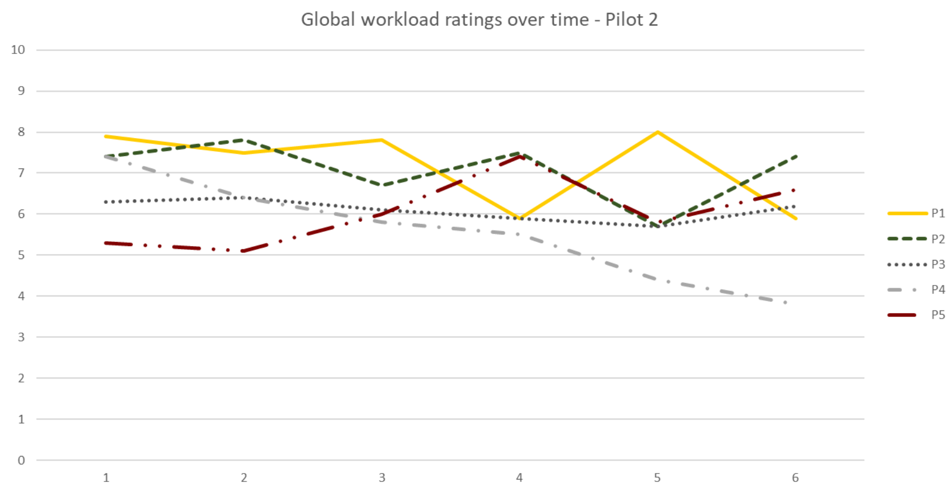


Figure 5.8: Global workload ratings from participants in Pilot 2

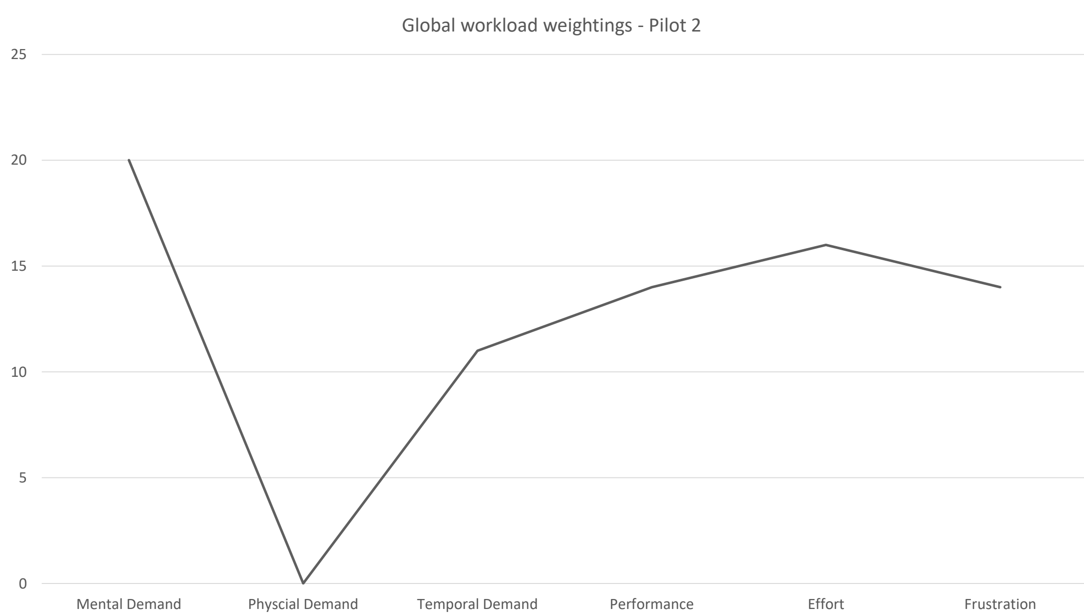


Figure 5.9: The weighting given to each workload measure by participants in Pilot 2

5.5.3 Task Performance

No explicit performance measures were collected during Pilot 1 as this pilot test was primarily concerned with exploring *if* and *when* participants used the decision support tool. However, several participants raised queries about the purpose of their classifications and the consequences should they make errors. Therefore, to increase the realism of the task the purpose of the classification decisions participants will be asked to make will be made explicit during the training session. This will be achieved by including an immersion aspect to the training which will set the context of operations on board a RN vessel. Additionally, participants will be made aware of the consequences of making classification errors, i.e. an incorrect decision could result in a fratricide incident.

There were 142 decisions that needed to be made to classify all tracks over the course of the scenario in Pilot 2. Correct classifications made by participants ranged from 40 - 89. Table 5.7 depicts the total number of classification decisions made by each participant, the number of correct decisions made, the number of errors and the number of omitted decisions. These descriptives show the challenging nature of the task with all participants making errors and failing to correctly classify all tracks.

Table 5.7: Descriptives of decisions made during Pilot 2

ID No.	Number of total decisions made	Number of errors made	Number of omitted decisions	Number of classifications automated system made
1	66	15	46	20
2	58	22	54	10
3**	95	6	17	10
4**	89	14	23	0
5**	56	16	56	20

** = had prior experience with the task and system

5.5.4 Experience

During Pilot 2, the two participants who had not taken part in the first pilot test utilised the DSS in phases 4 and 5 of the experiment compared to the participants who had taken part in the first pilot. As one participant wrote in the qualitative feedback following the task: *"After the first 10 second burst of the DSS and noticing that it did not perform as many classifications as I had anticipated therefore I was saving the second DSS usage 'for a rainy day' where the workload increased which didn't happen again during the remainder of the experiment"*. From prior experience on the task this participant decided to use the support tool only when they knew they could not complete the required number of classification decisions themselves in the time allocated. Supporting the literature that posits that confidence in ones' own ability to complete a task is a factor behind deciding to use an automated system or not (Lee and Moray, 1992).

It is possible that the two novice participants utilised the tool during the late phases as they spent phases 1-3 trying to get to grips with understanding how to do the task manually. One participant when asked why they did not choose to use the system at an earlier point responded, *"[I] was too busy trying to classify I did not think of it"*.

Although there were 11 weeks between Pilot 1 and Pilot 2 the results tentatively suggest that experience may lead to improved performance. Two of the three participants who performed in both pilot tests showed the best performance making 79 and 75 correct decisions in Pilot 2. Therefore, it may be expected to see an overall improvement in participant performance between the two-time trials of the main experiment.

5.6 Discussion

This chapter has discussed the development and initial testing of ARCS. ARCS has been designed to incorporate adaptable functions to provide researchers with a flexible micro-world. This flexibility facilitates research into how individuals interact with an automated system in near to operational environments. Although the limited numbers of participants at both pilot studies restricts the statistical analysis of the data collected, the main aim of testing the utility of ARCS was achieved. The majority of participants found the scenario challenging enough to warrant using the automated system and they provided a range of rationales for their decision to use the automated system or not. The elicitation of these rationales is a crucial, and yet often overlooked feature of exploring how individuals interact with automated systems. ARCS has been developed alongside a questionnaire based measure to ensure that these rationales are captured and incorporated into the analysis to provide contextual richness to any statistical findings.

Using immersive simulation environments increases the ecological validity of the research whist also maintaining experimental control (Brehmer and Dörner, 1993). The development of simulated computer environments in academia can provide an alternative way in which new systems can be developed and tested in collaboration with serving personnel. Although industrial partners do currently have the capabilities to develop and maintain simulated computer environments, for example, a complete replica of the Combat Management System (CMS) in use by RN personnel today, there are often competition and financial barriers to the collaborative use of these test environments between industrial and academic partners. Highlighted by the findings from the previous chapter, open architecture training environments are required and are paramount to developing automated systems which support operations in the field. Academia is uniquely placed to provide open architecture environments that are designed to promote collaboration and knowledge sharing as academic institutions are less constrained by commercial competition, ARCS is one such environment.

5.6.1 Recommended task changes/ARCS changes

From the two pilot tests several recommendations were highlighted to further increase the immersion of the task and the likeness to reality. Each recommendation will be presented below alongside how it was addressed for the main experiment:

(i) *Increase the task complexity* - the number of tracks presented to participants was increased from 53 in Pilot 1 to 142 in Pilot 2 and will be over 200 in the main experiment. Additionally, the number of hostile tracks will be raised from 8 to 10 during the critical time point. These changes will increase the complexity of the task and should be reflected in the workload scores provided by participants during the experiment.

(ii) *Provide participants with a longer, more detailed brief prior to the training session* - A short immersion feature will be built into the training session to aid the psychological immersion of participants into the task, increasing the realism of the task. Further information relating to the system itself will be provided to participants to mitigate against participants making appraisal errors. All slides included in the training brief can be found in Appendix Four.

(iii) *Include graphical icons for the locations of airport and bases* - introduction of further graphical features onto the main map will further support the realism of the task. Additionally, the graphical icons will provide participants with more information points with which to make their classification decisions.

(iv) *Provide participants with a greater number of times they can select the automated system* - in response to participants first use of the system to test its capability it was decided that participants will be able to select the system three times in the main experiment. Three was decided upon as it is expected that participants will use the system the first time to see how it operates. The second two selections, if chosen, may then highlight additional motivations behind automation uptake, for example as a way to reduce high levels of workload.

(v) *Questionnaire adapted to electronic version* - to improve ease of completion and prevent missing data the questionnaire given to participants will be transferred to

Qualtrics. Participants will be told to follow a link and complete the questionnaire online after they have completed the main task.

5.6.2 Conclusions of the development of ARCS and the benefit to research

The development of ARCS is one of the ways in which new automated decision support systems can be tested in near-to-real conditions. The flexibility of the simulation enables the replication of real-life challenges and conditions in which to robustly test the capabilities of new tools. The use of high-fidelity simulation testbeds and microworlds in the validation of newly developed automated tools and support systems is an underutilised methodology. However is an approach that could hold the key to how tools are transferred from sterile research environments into the 'real-world' without falling into the camp of 'brittle technology' (Woods, 2016). Additionally, developing highly realistic test-bed environments will support collaboration between personnel and system designers. The next chapter will present the findings from the full ARCS experiment.

Chapter 6

Using ARCS to explore automation usage decisions

6.1 Abstract

This chapter presents the Automatic Radar Classification Simulation (ARCS) experiment which aimed to explore the motivational and cognitive factors that may influence uptake of an automated decision support system. Students from the University of Liverpool (N=42) took part in two trials of the ARCS microworld experiment in return for course credit or reimbursement. Interesting results were observed as a high percentage of participants selected to use the automated system at both trials, and training effects were observed with task performance improving across all conditions at trial 2. Of particular interest is the finding that there were no significant differences in task performance with or without access to the decision support system. Similarly, workload was not quantitatively found to influence the uptake of the decision support system but was qualitatively referred to by participants as a factor in their decision-making process. Additionally, no clear associations between several cognitive traits and uptake of the automated system were observed. The findings from this experiment have several interesting implications in relation to the importance of training, transparency of system functionality and the need to maintain a sense of agency over task completion.

6.2 Introduction

This chapter explores how several factors, motivational and cognitive, interact with the decision to use an automated system or not and task performance. This chapter presents the findings from the student experiments conducted with ARCS and therefore will:

(i) Summarise how automation use can impact on task performance, in terms of the influence of:

  (a) Information provided about the automated system

(b) Experience of task and system

(ii) Describe how automation uptake and use can be influenced by two motivational factors:

(a) Workload

(b) Accountability

(iii) Describe how individual differences in cognitive processing styles may modulate automation uptake and use through discussing six cognitive processing styles:

(a) Cognitive Flexibility

(b) Need for closure

(c) Self-regulation processes of Assessment and Locomotion

(d) Impulsivity

(e) Conscientiousness

(f) Propensity to Trust

(iv) Present findings from the ARCS mixed-design experiment conducted with students

(v) Discuss implications of the findings

6.2.1 How automation use can impact on task performance

Research has highlighted the benefits automation use can have upon task performance across a variety of domains (Moray et al., 2000; Parasuraman et al., 2012; Schraagen and van de Ven, 2008; St. John et al., 2005). However, automated systems can result in derailments of performance such as automation complacency (Beck et al., 2007) and loss of situational awareness (Endsley, 1987), see Chapter 2. It is interesting that automation use is often posited as the solution to dealing with increasingly complicated and dynamic military environments. Yet, as discussed in Chapter 4, the application of automated systems into operational environments does not necessarily result in immediate improvements to task performance, especially if the tools are not fit for purpose.

The introduction to this thesis discussed the crucial factor of system reliability and the impact this has upon trust in a system. Trust is commonly linked to system use (Lee and See, 2004) and research has suggested that providing users with meta-information of the automated system can improve trust calibration (McGuirl and Sarter, 2006; Seong and Bisantz, 2008; Wang et al., 2009). This is due to developing the users' cognitive understanding of the system; both in terms of what it can and can't do and why. Lee and Moray (1992) posit that trust towards machines is transient over time and heavily influenced by the reliability of the system, therefore experience of the system will influence current trust of that system. Training also aids the development of learned trust (Hoff and Bashir, 2015) which facilitates better calibration of trust towards the automated system thereby supporting appropriate automation usage decisions. This assertion was supported by Beck et al. (2007) who found that with feedback and training on a target detection task, disuse of the automated system reduced to 27% as opposed to the 55% that was observed without training and feedback. Trust in a system has also been associated to the anthropomorphic features of the system (Hancock et al., 2011). In fact, individuals have a tendency to anthropomorphise technology even if no human-like features are built into the system (Nass et al., 1995), suggesting a desire to comprehend and approach working with a system in the same way as we comprehend working with other people.

A crucial factor in taking a decision-centred approach to designing automated systems is instilling a sense of agency in the operator. Forming a fundamental aspect of an individual's self-awareness (Gallagher, 2002), a sense of agency promotes self-efficacy. A recent review by Limerick et al. (2014) explored the limited number of studies that have explicitly looked at the ways in which sense of agency was influenced by interactions with automation. As expected, the Level of Automation (LOA) directly impacts upon an individual's locus of control. That is, as the LOA increases the perceived control over the task declines. Limerick et al. (2014) argue that identifying the tipping point to when users no longer feel a sense of agency over the task will inform future design and development of collaborative systems. This concept is crucial to ensure that human-machine interaction and collaboration functions as an effective team dyad. Bekier et al. (2012) explored this tipping point for air traffic controllers, collecting 500 responses to an on-line survey. Their findings argue the importance for locus of control to remain with the operator, as air traffic controllers cited a shift away from utilising automated tools when they perceived a shift in locus of control away from themselves. Further, Kaplan et al. (2001) found that the illusion of control created by task involvement lead to increased reliance on a statistically valid decision aid to complete financial prediction decisions. It should be noted however that the preferred aid of the Subject Matter Experts (SME) (N=91) and business masters' students (N=61) who took part in this study was not the most useful aid but the more interactive one.

Drawing also from the organisational and social psychology domains, the link between perceived controllability over a task and effort has been well established; the lower the perceived controllability the lower the effort (Litt, 1988). Perceiving organisational functioning as influenceable, promotes resiliency of self-efficacy (Bandura and Wood, 1989), self-efficacy in turn promotes the likelihood of task success. Additionally, feelings of investment in a task (i.e. goal setting) can facilitate task completion (Yearta et al., 1995) therefore, allowing operators to decide when, and if, to use an automated system may increase their perceived control and investment in the task as they have the final decision on how to approach completing that task. Maintaining a sense of agency over the task may also reduce the occurrence of automation complacency errors. If the operator remains engaged in the task, due to perceiving the importance of their role, that engagement may mitigate against automation complacency, via maintained effort.

How confident an individual is in their own skill at a task has also been associated with automation uptake and use (Lee and Moray, 1992). Individuals compare perceptions of their own abilities to the perceived utility of the decision aid (Powell, 1991). When trust in the system is greater than an individuals' self-confidence in their own skill, then the automated system is used (Lewandowsky et al., 2000). Conversely, if an individual is confident in their ability to complete the task manually they will be less inclined to utilise a decision support tool. Therefore, it is hypothesised that,

- H1: Individuals with access to the Decision Support System (DSS) during the task (Groups B & C) will perform better (i.e. correctly classify more tracks-classify all hostile tracks) compared to the control group who do not have access to the DSS.

- H2: Individuals who are provided with information that describes the automated system in relation to a human operator (low automation condition) will select the DSS more times than individuals who are provided with information that describes the automated system in relation to current technological capabilities (high automation condition).

6.2.2 Motivational influences on automation usage decisions

The most commonly explored performance metric when looking at the influence automated systems can have upon task performance is workload. In general, the application of automated systems has been found to decrease operator workload (for example see Balfe et al. (2015); Röttger et al. (2009)). However, automation applied to the stages of information analysis and decision making has been found to degrade performance and increase workload (Kaber et al., 2005). Additionally, in operational settings practitioners have reported that workload can be increased when utilising automated systems (see findings from Chapter 4). If using an automated system increases workload the system will be disused (Parasuraman and Riley, 1997), it therefore remains of interest for researchers to understand the impact automated systems have upon operator workload.

A second user assessment measure that has been linked to automation usage decisions is accountability. Accountability can lead to avoidance of responsibility for decisions (Eyre et al., 2008a) and is posited to influence an operator's motivation to utilise an automated system. The legal precedents for the use of automated weapons and systems in warfare are still being established as technology is continuing to break boundaries; changing the way in which warfare is fundamentally fought. Currently, the main concern from an international law perspective, and ethically, is whether the rules of distinction and discrimination are still followed when utilising advanced warfare capabilities, such as remote attack systems (Boothby, 2014). The principal concern is who is liable if an error were to be made: would the operator be personally liable for the erroneous attack even though they are technically miles away from the impact site?

In an experimental navigation and gage monitoring task, accountability was manipulated to explore the impact perceived accountably had upon the occurrence of commission and omission errors of automation use. Skitka et al. (2000) manipulated accountability by informing participants that their performance was being recorded and they would be required to explain their strategies used to meet the task performance requirements. They found that increased personal accountability reduced automation bias, via increasing verification behaviour. Moreover, in an industrial plant task experiment it was found that participants who made task errors when cooperating with a computer experienced reduced self-confidence, however, in the same task when cooperating with two other individuals, task errors did not influence self-confidence (Lewandowsky et al., 2000). The authors posit that a diffusion of responsibility occurred when cooperating with other individuals but not when cooperating with a computer. What this means is that individuals were holding themselves more accountable when working with an automated system as opposed to when they were working with other people. The presented research above used experimental designs were automation was incorporated into the task as opposed to allowing participants the choice of using the system or completing the task manually. To the authors knowledge, at the time of writing this thesis, no research has been conducted to explore the possible influence accountability has upon the choice to use an automated system. Considering this literature this study hypothesises that,

- H3: Individuals will use DSS when task demands exceed cognitive capacity (when workload is high).
- H4: Individuals in low accountability condition will select DSS (more times) compared to individuals in high accountability condition.

6.2.3 Cognitive processing styles and how they may modulate automation uptake and use

Historically research on human-machine-interaction has focused on the attributes of the automated system and how they impact upon use. However, recently there has been a move towards focusing on features of the human operators of these systems. Does a relationship between certain cognitive processing styles and automation use exist? There is limited literature from the human-machine-interaction domain, however there is a wealth of research that has been conducted exploring cognitive processing styles and decision making in uncertain, time pressured and critical environments (see Chapter 2). The use of automation is posited to aid human operations in such environments therefore Naturalistic Decision Making (NDM) literature will be drawn from to explore the potential relationships between six cognitive processing styles and automation use.

Cognitive Flexibility Inventory

Defined as *"the ability to switch cognitive sets to adapt to changing environmental stimuli"* (Dennis and Vander Wal, 2010, p. 242), Cognitive Flexibility (CF) is a cognitive processing style that has been linked to the ability to make decisions under uncertainty. CF has also been associated with supporting adaptive expertise, which is argued to be crucial in sociotechnical environments (Hoffman et al., 2013). CF is measured along two subscales, CF-control and CF-alternatives. CF- control is associated with a disposition towards perceiving challenging situations as controllable and has been associated with reducing task based uncertainty (Power, 2015). Whereas CF- alternatives has been related to the tendency to perceive multiple solutions to difficult problems (Dennis and Vander Wal, 2010) and has been associated with increasing outcome based uncertainty (Power, 2015). It is therefore anticipated that individuals high on CF will utilise the decision support system later compared to individuals scoring low on this measure. This is due to individuals high on CF being adaptive when processing complex environments (Martin and Anderson, 1998).

Need for closure

The Need for closure (NFC) scale assesses the trait desire towards definite answers as opposed to ambiguity or confusion (Kruglanski, 1989) and has been found to affect the extent of information searching and processing required to reach a judgement. That is to say, contextual factors such as time pressures or mental fatigue can lead to extended information searching to be viewed subjectively as costly, therefore this stage of decision making is swiftly closed in order to reach a decision and gain closure on this subject and/or task (Jost et al., 2003). Therefore, Kruglanski et al. (2010) highlight that NFC underpins the tendency to *'seize'* on early information/evidence and *'freeze'* upon the decision such evidence supports. Accordingly, it is anticipated that individuals scoring high on NFC will utilise the decision support system earlier, as a consequence of subjective workload, in order to facilitate the rapid closure of the decision task.

Assessment & Locomotion

Classically conceived as aspects of the self-regulatory system, assessment and locomotion are involved in the ability to desire, achieve and maintain end states and/or goals. Assessment encompasses the analytical process of critically evaluating states, goals and means to judge relative quality, whereas locomotion concerns the movement between states (Kruglanski et al., 2007). Therefore, it is anticipated that individuals high (vs. low) on locomotion will utilise the decision support system earlier in the task. Not only in order to achieve the desired state of movement towards a new state (i.e. task completion) but also due to possessing greater flexibility towards task changes. Whereas, individuals high (vs. low) on assessment are anticipated to utilise the decision support system later in the task due to focusing greater attention on analytically understanding the situation and task itself before making the decision to use the tool or complete the task manually.

Impulsivity

The Barrett Impulsivity Scale (BIS)-11 is the most commonly used assessment for impulsivity in both research and clinical settings. The scale measures 3 aspects of impulsiveness: (i) Attentional impulsiveness - concerns the ability or inability to focus or concentrate attention; (ii) Motor impulsiveness - concerns acting without thinking; and (iii) Non-planning impulsiveness - concerns the tendency to lack forethought or plan for the future (Barratt, 1985). Subtypes of impulsivity are related to different aspects of executive control of working memory (Whitney et al., 2004). Although a complex relationship, it has been posited that impulsivity may be displayed due to individuals lacking working memory resources which prevents them from assessing multiple options when faced with complex decisions (Hinson et al., 2002). Impulsivity has also been linked to attention lapses and/or distractibility (Levine et al., 2007), which could have a detrimental impact on task performance. Accordingly, it is anticipated that individuals scoring high on the BIS are more likely to utilise the automated tool early in the task compared to individuals who score low on the BIS. Additionally, individuals who score highly on attentional impulsiveness scale may perform more poorly compared to individuals who score low on this subscale.

Conscientiousness

Conscientiousness, agreeableness, neuroticism, extraversion and openness comprise the Five Factor model of personality. Specifically, conscientiousness is concerned with an individual's active self-regulation of behaviour; primarily to achieve goals, organise, plan and complete tasks (Costa Jr and McCrae, 1992). In a threat detection study where participants were required to identify if an image (or block of images) contained a hostile person or not conscientiousness was found to be positively associated with task performance (Szalma and Taylor, 2011). Further, Barrick and Mount (1991) found that conscientiousness was one of the strongest predictors of job performance. It has been posited that individuals who score high on conscientiousness would be less susceptible to automation complacency as they will continue to engage in order to achieve their task goals (Szalma and Taylor, 2011). However, counterintuitively high personal investment in a task has been linked to increased disuse of an automated system, even though the participants were aware that the system outperformed them in the task (Beck et al., 2009). These findings suggest that a tipping point may exist with enough conscientiousness, and therefore investment in a task, facilitating the appropriate application of an automated system to help achieve task goals. However, too high a level of conscientiousness, and therefore investment in a task, may link into perceived accountability and therefore result in disuse of the automated system.

Propensity to trust

The propensity to trust scale was designed to gain a measure of how positive one feels towards automation (Merritt, 2011) and has been associated with patterns of trust towards automated systems over time. Merritt et al. (2009) found that when operating with a system that had transparent functionality, propensity to trust scores were significantly associated to trust in the system, however if the system functionality was ambiguous implicit attitudes towards automation were associated with trust in that system. To date no study has explored if high scores on the propensity to trust scale are associated to system use in a high-fidelity maritime scenario.

6.2.4 Summary

Sociotechnical systems are influenced by a number of factors and research has shown that the application of automated systems can support and/or hinder task performance. The literature highlights that a range of factors feed into what underlies effective use of automated systems and equally can result in the overuse or complete disuse of these systems. However, there has been a tendency to focus specifically on features of the system and not on the cognitive traits of the human operators. Therefore, it is important to extend the understanding into human-machine-interaction by exploring the cognitive traits that have yet to be looked at in relation to their potential influence behind automation usage decisions. Therefore, the final two hypotheses addressed in this chapter are:

- H5: To explore the relationship between scores on each cognitive trait and automation use.
- H6: To explore the relationship between scores on each cognitive trait and task performance.

6.3 Method

In total N=42 participants took part in the experimental task. All participants were undergraduate, masters, PhD or postdoctoral researchers from the University of Liverpool. Participants were split into three conditions: A (n=13), B (n=13), and C (n=12). Three participants, two from condition B and one from condition C, dropped out. One participant from condition A was removed due to technical problems with the task at trial 1. This resulted in a final sample of N=38; 15 males and 23 females.

A non-significant chi squared analysis showed there was no difference in gender between the conditions ($X^2(2)$= 6.877, p =.032). A one way analysis of variance revealed a significant difference between age and condition (F(2,35) = 11.378, p = .000). Post hoc tests using the Bonferroni correction revealed a significant difference between conditions A (M = 24.23, SD = 1.79) and B (M = 19.62, SD = 1.98), p=.000 and conditions B and C (M = 23.83, SD = 3.97), p = .001. However, no statistically significant difference was found between conditions A and C (p = 1.000). Table 6.1 Provides full breakdown of the participant demographics. The study received ethical approval from the University of Liverpool and MoDREC (Protocol No: 785/MoDREC/16).

Table 6.1: Demographic breakdown of age of participants across conditions

		Condition		
		A	B	C
N		13	13	12
Gender	Male	5	2	8
	Female	8	11	4
Age	Mean ±SD	24.23 (±1.79)	19.62 (±1.98)	23.83 (±3.97)
	Range (yrs)	22-28	18-24	18.29
	Median	24	19	24.50

Due to the small sample size, median scores are provided alongside mean and standard deviations

6.3.1 Experimental Procedure

A mixed design was used (see Table 6.2) with participants taking part in the experiment in groups (5 or more, depending on practicality and availability of participants) to replicate working with other personnel around. Each participant took part in the study twice, with at least a 2 week break in-between experiments. Participants were allocated into one of three groups (A, B or C), they completed the experiment within the same automation background condition, and experienced both the high and low accountability conditions. The accountability condition was therefore counterbalanced, half the group experienced low accountability at study time 1 and high accountability at study time 2. For the other half of the group this was reversed- high accountability at study time 1 and low accountability at study time 2. The accountability prime was a paper based inject that was given to participants following the training presentation and the automation background conditions divided participants based on the information they received about the automated system, all materials used can be found in Appendices Three - Six.

Prior to data collection an ideal sample size was computed using GPower 3.1. (Faul et al., 2007) with parameters of a medium effect size (0.25), p=0.05 and alpha level 0.8. This power calculation led to a sample size of 69 participants being required to ensure appropriate experimental power. However, due to the design of the experiment and the length of time it took to complete the task a final sample of N=38 were recruited, the limitations of this are discussed.

Table 6.2: Experimental Design

		Automation Background (between-participants)		
		Control (A)	Low (B)	High (C)
Accountability (within-participants)	Low ($_L$)	A_L (*n=13*)	B_L (*n=13*)	C_L (*n=12*)
	High ($_H$)	A_H (*n=13*)	B_H (*n=13*)	C_H (*n=12*)

6.3.2 The Training Session

Table 6.3 provides a brief overview of the stages to the training session (all material used in study 1 and study 2 is included in Appendices Three, Four and Five). Following the training session participants were briefly reminded of their task instructions and handed the paper based accountability prime before beginning the main session. The second time participants completed the task the immersion videos were not played again but were discussed alongside stills taken from the videos. The 15-minute training session for participants to understand where the information would be on the screen and how to complete the task was also optional. These slight changes to the training session for the second trial were deemed a way to mitigate against participants cognitive disengaging from the task prior to the main scenario as the information had been covered previously.

6.3.3 The Main Task

The task of the participant was to classify each radar track into hostile, neutral or friendly. To complete the experiment, the participant must remain vigilant to the radar tracks currently on screen, ensuring that all tracks are correctly classified and any radar tracks that are deemed hostile are highlighted to command. For the main task scenario, conditions B and C had the option to defer to the automated tool if they felt it was appropriate. Participants could defer to the system three times throughout the 30-minute scenario. Participants in condition A were the control group who did not have access to the automated system, these participants were required to complete the task manually.

The main task ran for 30 minutes, see Figure 6.1. Participants were tasked with classifying all the tracks on the screen as accurately as possible. At 10-minute intervals, the radar screen froze and participants completed NASA TLX ratings (see Appendix Four for the NASA TLX measure). This was intended to provide a continued gauge of participant workload throughout the task. Two scenarios were developed as participants would take part in the experiment twice, Figure 6.2 and Figure 6.3 display the total number of tracks on screen for each scenario. For scenario 1 participants had 207 tracks to classify and for scenario 2 there were 209 tracks to classify. Presentation of the scenarios was counterbalanced across conditions. To replicate daily operations on a ship it was important that both scenarios were of similar difficulty to simulate the day to day situations operators face, a Fisher's exact test confirmed that workload was found to not significantly differ between scenarios 1 and 2 at trial 1, $p = .495$ or trial 2, $p = 1.000$ suggesting that both scenarios were of a similar difficulty.

6.3.4 The Questionnaire Battery

Having completed the scenario participants were asked to complete the questionnaire battery (see Table 6.4). This consisted of 7 sections (A - H), including demographic questions, qualitative feedback questions (see Table 6.5) and the six cognitive trait questionnaires to explore if a relationship between these traits and automation usage decisions exists. The control group were only shown questions 1, 2, 7 and 8 of Table 6.5. The complete questionnaire battery is provided in Appendix Five.

Following completion of both trials participants received course credit or £10 for their time.

Table 6.3: The stages of the training brief given to participants

	Training procedure for ARCS
1	Participants were immersed in the task via video and audio footage of when high profile erroneous automated usage decisions have been made in recent history. Additionally, a short video was played to participants to inform them of the operational environment the air defence task is conducted in.
2	The main task was then explained to participants, highlighting the requirement for their classification decisions to be as accurate as possible. The division of the experimental task was also explained to participants, that they would first have 15 minutes to get used to the interface and how the classification process worked before beginning the main 30 minute task. Participants were shown what the interface looks like, where the key information is located and how to interact with the simulation.
3	Next, participants were shown the iconography they can expect to see throughout the task. For example, what airports will look like on the radar picture. It is at this stage that participants are asked to complete the symbology sheet in front of them, allowing them to label each symbol. This symbology sheet remained with them throughout the task-forming a reference point for the symbology on screen.
4	The participants were then provided with examples of behaviour that they might expect to see during the task. A table of given speeds and altitudes was provided to each participant, again as a reference sheet they can use throughout the task. As the sample population are novices to the air defence task, certain behaviours that would infer a potentially hostile track were explained.
5	Following this, the NASA TLX questions were explained to participants. Definitions of each of the NASA TLX workloads were also given as a reference sheet.
6	For the automation conditions (conditions B & C) the characteristics of the automated decision support system were explained. For instance, that participants could use the system three times (for 10 seconds at a time) throughout the 30 minutes. Participants in condition B were informed that *"The tool is as reliable as a well-trained experienced operator"*, and condition C were informed that *"The tool is as reliable as current technology can be"*. This difference in information was used to explore if a difference in automation use would be observed depending on what information was provided to participants about the system.
7	Finally, participants were asked to read the 'Daily Tasking Orders' in front of them. These orders reminded participants on their task and were also the inject for manipulating accountability.

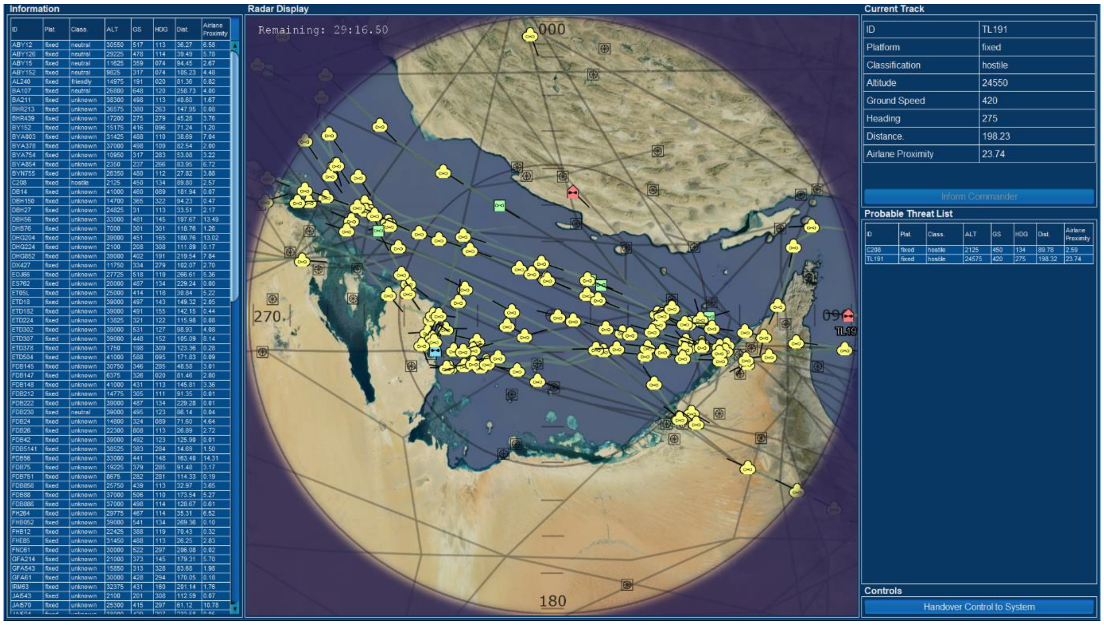


Figure 6.1: Screenshot of the main task screen within ARCS

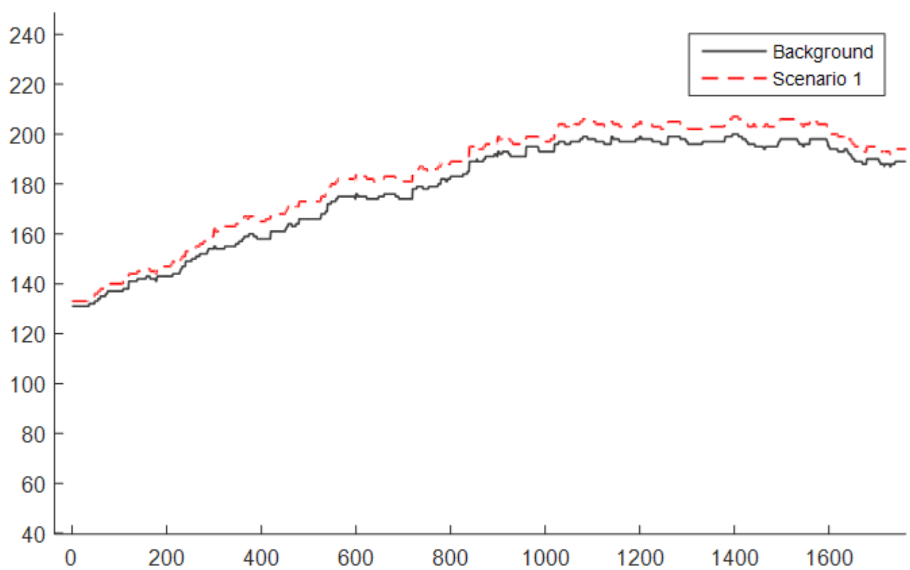


Figure 6.2: Number of tracks on screen per second for scenario 1

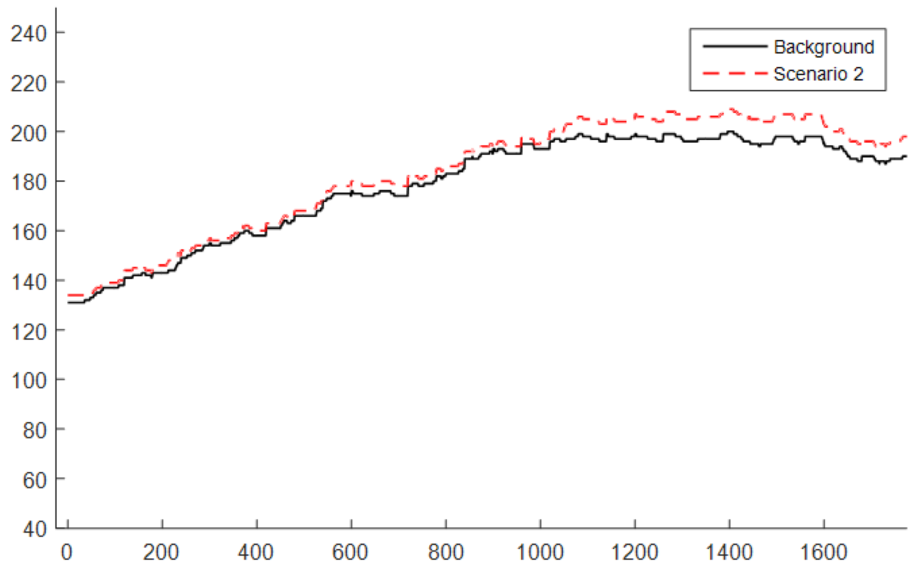


Figure 6.3: Number of tracks on screen per second for scenario 2

Table 6.4: Overview of the sections included in the questionnaire battery at Stage 2

Section A	Demographics
Section B	Qualitative feedback questions
Section C	Cognitive Flexibility Inventory (Dennis and Vander Wal (2010))
Section D	Need for Closure (Kruglanski et al. (1993))
Section E	Assessment & Locomotion (Kruglanski et al. (2000))
Section F	BIS-11 (Patton et al. (1995))
Section G	Conscientiousness Scale (John and Srivastava (1999))
Section H	Propensity to Trust Scale (Merritt et al. (2013))

Table 6.5: Qualitative questions included in questionnaire battery

1	From 1-10 how would you rate your performance? (1= poor, 10= perfect)
2	On a scale of 1-10 how accountable did you feel towards the decisions you made? (1=not accountable, 10= very accountable)
3	Did you choose to use the decision support system at any point(s) during the scenario? (YES/NO) (a) Why did you choose to use the decision support system?
4	Reflecting on your performance, where there any occasions that you think you should have used the decision support system but did not and why?
5	Reflecting on the scenario, where there any occasions that you felt you could have use the decision support system earlier but did not?
6	What made you not choose the system at an earlier point?
7	Did you recheck tracks once you made a classification decision? (a) If Yes, how often did you recheck and why?
8	If you could perform the scenario again, is there anything you would do differently?

6.4 Results

6.4.1 Summary

The six hypotheses and summary findings are outlined in Table 6.6. Opposing the hypothesised benefit of having access to the DSS no significant differences were found between the conditions on the performance metrics collected. No associations were found between subjective feelings of workload, accountability and the information participants received about the DSS and their subsequent selection to use the system or not. However, it was shown that participants across all conditions significantly improved at the task during trial 2, reflected in the total number of correct decisions made, the reduction in incorrect decisions and lower subjective workload ratings. No clear interactions between the cognitive traits measured and system use or task performance were observed. However, interesting lines for future research are discussed and these findings have interesting implications for further research into perceived performance benefits of using automated systems.

A vast amount of data was collected by ARCS and only analysis relating to the six hypotheses presented in Table 6.6 is included in this chapter. Appendix Seven provides examples of the ways in which analysis can be conducted on the data ARCS collects.

Table 6.6: Summary table of findings from hypotheses

Hypothesis	**Findings**
H1: Individuals with access to DSS during task (Groups B & C) will perform better (i.e. correctly classify more tracks- classify all hostile tracks) compared to control group	Contrary to H1 it was found that there were no significant differences between conditions. Whilst not significant the control group overall made fewer mistakes and more correct decisions compared to the conditions with access to the DSS.
H2: Individuals in low automation condition will select DSS more than individuals in the high automation conditions	No significant differences were found between automation conditions and selection of DSS.
H3: Individuals will use DSS when task demands exceed cognitive capacity (when workload is high)	Workload was not related to participants selection of DSS. However, workload was found to significantly reduce at trial 2 compared to levels at trial 1 across all conditions.
H4: Individuals in low accountability condition will select DSS (more times) compared to individuals in high accountability condition	Accountability was found to be an internally generated construct as opposed to being externally manipulated and was not related to participants selection of DSS.
H5: Explore the relationship between scores on each cognitive trait and automation use	No clear associations between scores on each cognitive trait and automation use were found.
H6: Explore the relationship between scores on each cognitive trait and task performance	No clear associations between scores on each cognitive trait and task performance were found.

6.4.2 H1: Individuals with access to DSS during task (Groups B & C) will perform better (i.e. correctly classify more tracks- classify all hostile tracks) compared to control group

Total correct classifications made

A 2x3 mixed ANOVA was conducted to investigate the impact of participant condition on the number of correct classification decisions made. There were no significant outliers, data was normally distributed and homogeneity of variance was present as assessed by Levene's test of homogeneity of variance ($p > .05$). There was also homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p > .910$). Mauchly's test of sphericity was not required as there were only two within-subject factors. There was no statistically significant interaction between condition and the number of correct decisions made, $F(2,35) = 1.448$, $p = .249$, partial $\eta^2 = .076$. The main effect of condition was found to be not statistically significant, $F(2,35) = 1.870$, $p = .169$, $\eta 2 = .097$. However, a main effect of trial showed a statistically significant difference in mean total correct classifications at trial 1 and trial 2, $F(1,35) = 19.663$, $p = .001$, partial $\eta^2 = .360$. There was an increase in the number of correct classifications made at trial 2 (M = 194.79, SD = 9.84) compared to trial 1 (M = 142.03, SD = 11.28), this was a statistically significant mean difference of 52.76 correct decisions, 95% CI [28.61, 76.91], $p = .001$. Further, a Bayesian statistical analysis also revealed very strong evidence for the statistically important differences between groups (BF = 228.8). Therefore, we can accept the alternative hypothesis - there is a difference between the total number of correct classifications at each trial. This analysis is further supported by an increase in participants stating that they were happy with their performance following completing the task the second time around.

Total correct neutral classifications made

A 2x3 mixed ANOVA was performed and revealed a non- significant interaction between condition and trial on total number of correctly identified neutral tracks, $F(2,35) = 1.283$, $p = .290$, partial $\eta^2 =.068$. No significant main effect of condition was found, however, a significant main effect of trial was revealed, $F(1,35) = 18.524$, $p = .001$, partial $\eta^2 = .346$. Post-hoc tests with Bonferroni correction revealed that on average participants correctly classified 50.89 more neutral tracks during trial 2 compared to trial 1 , 95% CI [26.87, 74.86],$p = .001$.

Total correct friendly classifications made

A 2x3 mixed ANOVA was performed to investigate the impact of participant condition on the number of correctly classified friendly tracks. A significant interaction between condition and trial on total number of correctly identified friendly tracks, $F(2,35) = 6.176$, p = .005, partial η^2 =.261 was found. There was a statistically significant effect of trial on total correctly classified friendly tracks for the control group, $F(1,12)$= 12.844, p = .004, partial η^2 =.517, and for the high automation group $F(1,11)$= 26.400, p = .001, partial η^2 =.706, but not for the low automation group $F(1,12)$= .081, p = .781, partial η^2 =.007. Post-hoc tests with Bonferroni correction revealed that participants in the control group (M = 3.69, SE = 1.601, p = .020) and high automation group (M = 3.67, SE = 1.61, p = .001) significantly correctly classified more friendly tracks during trial 2 compared to trial 1. Table 6.7 displays the mean and standard deviations for each condition across both trials.

Table 6.7: Mean and standard deviations of correctly identified friendly tracks

	Trial 1		Trial 2	
Condition	M	SD	M	SD
Control (A)	2.38	1.19	3.69	1.60
Low Automation (B)	2.54	1.71	2.38	1.45
High Automation (C)	1.67	1.56	3.67	1.61

Total correct hostile classifications made

A 2x3 mixed ANOVA was performed, there was no statistically significant interaction between condition and trial $F(2,35)$= 1.290, p = .288, partial η^2 = .069. There was also no statistically significant main effect of trial $F(1, 35)$= 2.778, p = .105, partial η^2 = .074. Table 6.8 displays the means and standard deviations of the total number of correctly identified hostile tracks by each condition across both trials. No participant correctly identified all 10 tracks in both scenarios.

Table 6.8: Mean and standard deviations of correctly identified hostile tracks

	Trial 1		Trial 2	
Condition	M	SD	M	SD
Control (A)	5.85	3.87	5.77	3.11
Low Automation (B)	5.31	3.38	6.00	3.87
High Automation (C)	5.08	2.81	7.00	2.70

When looking at just the conditions with access to DSS, independent t-tests were performed to assess the differences between those who selected the system and those who did not and the total number of correct hostile decisions made. At trial 1 this was found to be not significant $t(23)$= .086, p = .932 and at trial 2 just significant $t(23)$= -2.083, p = .049. Participants who used the tool on average classified 7.21 hostile tracks correctly (SD = 3.03), compared to those who did not (M = 4.17 SD = 3.43). Further a Bayesian statistical analysis showed anecdotal evidence that there was a difference between groups (BF = 1.708). However, independent t-tests performed on the total number of correct neutral and friendly classifications made were all non-significant.

Correct decisions made across each 5-minute phase

ARCS recorded all performance metrics across the 30-minute task in 5-minute phases this allowed the breakdown of the performance data across each phase to enable a sequence analysis to be conducted. As the data was not normally distributed and contained several outliers (Figure 6.4), Friedman's test was computed to look at the differences between the distributions of correct decisions made by participants at each phase. A non-statistically significant difference between phase distributions was found at trial 1, $X(5) = 8.511$, $p = .130$. However, a statistically significant difference between the distributions of correct classification decisions between the 6 phases of the scenario was found at trial 2, $X(5) = 11.279$, $p = .046$, see Table 6.9. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons, with this correction significance was lost.

Table 6.9: Descriptives of correct classification decisions made each phase at trial 1 and trial 2

	Trial 1			Trial 2		
Phase	M	SD	Med	M	SD	Med
1	23.82	24.51	13.50	47.61	37.21	35.50
2	25.53	23.98	15.50	38.32	23.48	31.50
3	26.21	18.24	21.00	34.74	19.62	31.50
4	22.95	12.80	22.00	28.00	12.13	26.00
5	22.18	13.98	20.00	23.66	14.05	18.50
6	22.26	16.80	17.50	22.76	13.11	16.00

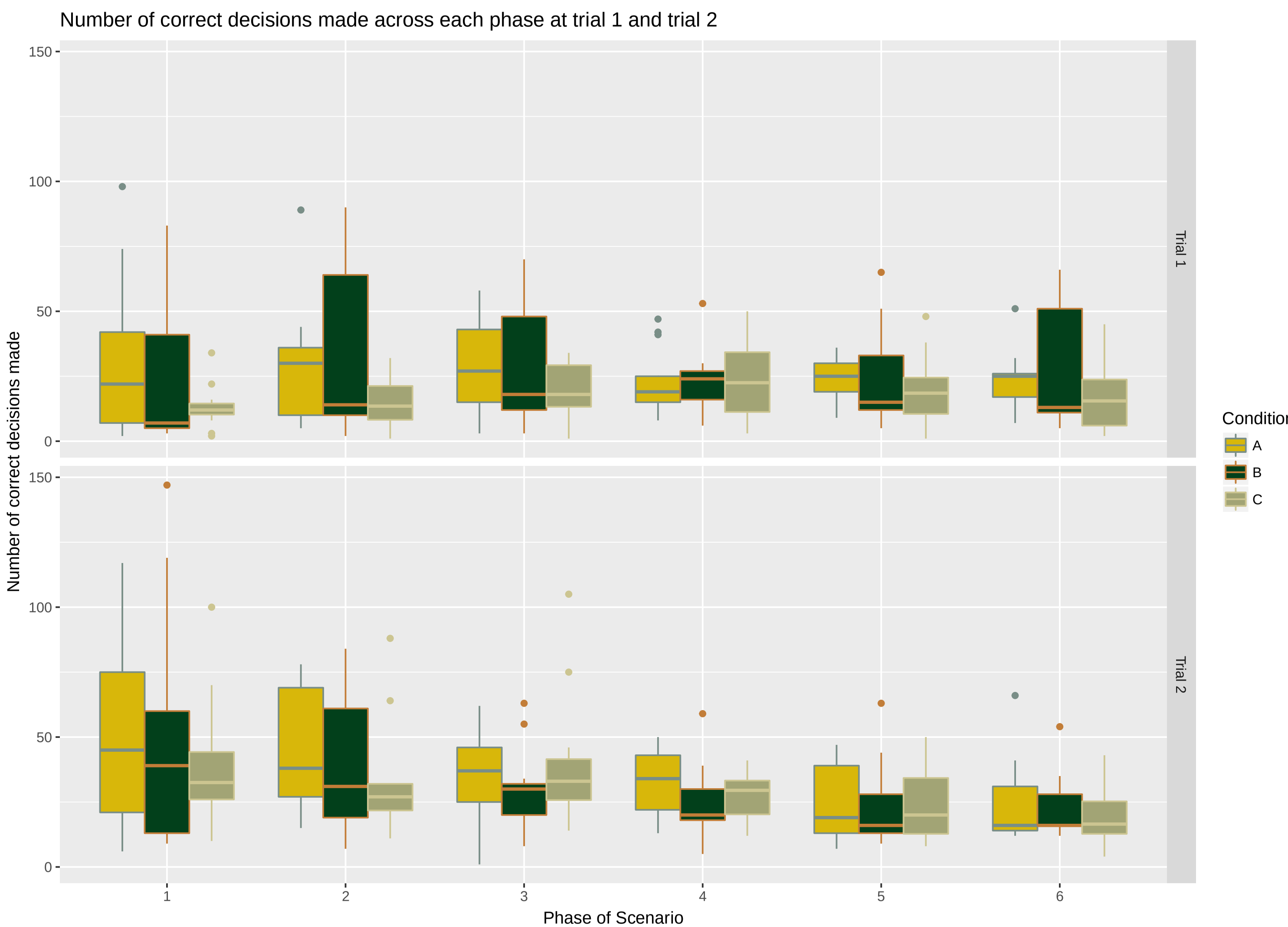


Figure 6.4: Boxplots of total number of correct decisions made each phase across trial 1 and trial 2

Incorrect decisions made across each 5-minute phase

A Friedman's test was computed due to the data violating assumptions of normality and containing several outliers (see Figure 6.5). The analysis revealed a statistically significant difference between the distribution of incorrect classification decisions made at each phase at trial 2, $X(5)= 12.412$, $p = .030$. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons and significance was lost. At trial 1 the analysis was found to be not significant, $X(5)= 7.297$, $p = .199$, see Table 6.10.

Table 6.10: Descriptives of incorrect classification decisions made each phase at trial 1 and trial 2

	Trial 1			Trial 2		
Phase	M	SD	Med	M	SD	Med
1	7.34	10.67	3.00	5.68	10.70	2.00
2	8.18	11.77	2.50	6.87	10.93	2.50
3	8.76	12.94	3.00	5.37	7.32	2.50
4	7.05	9.14	3.50	5.61	6.50	4.00
5	7.82	10.04	3.00	4.87	6.00	3.00
6	6.97	9.35	2.00	4.03	5.35	1.50

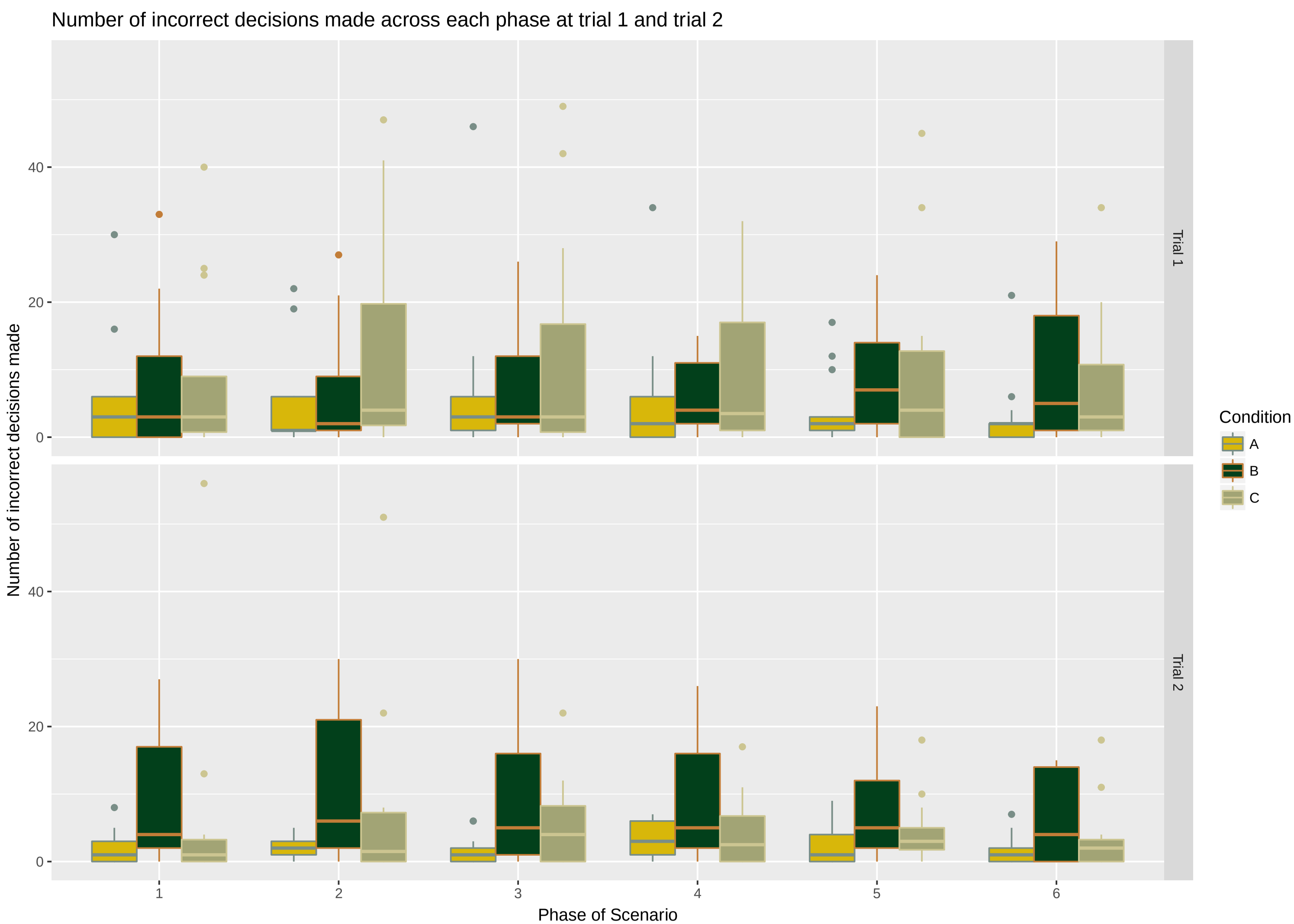


Figure 6.5: Boxplots of total number of incorrect decisions made each phase across trial 1 and trial 2

Associations between number of changed decisions and task performance

Pearson's correlation showed a significant positive association between the total number of correct changes to decisions and overall total number of correct decisions at trial 1, r = .457, p = .004, however significance was not found at trial 2, r = .229, p = .167.

Pearson's correlation showed no significant association between the total number of incorrect changes to decisions and overall total number of incorrect decisions at trial 1, r = .205, p = .217, however, significance was found at trial 2, r = .600, p = .001.

6.4.3 H2: Individuals in low automation condition will select DSS more than individuals in the high automation conditions

Chi-squared test for association was conducted between condition and selection of DSS. Two cell frequencies were lower than 5 therefore Fisher's exact test was read and found to be not significant at trial 1 (p = .550) or trial 2 (p = .363). Suggesting that there was no association between participant condition and selection of DSS (see Table 6.11).

Table 6.11: Percentage of participants who selected to use the DSS across trials 1 and 2

Condition	trial 1	Trial 2
Low Automation (B)	69% (n = 9)	69% (n = 9)
High Automation (C)	75% (n = 9)	83.3% (n = 10)

Selection of DSS at trial 1 and trial 2

Looking at only the data from conditions with access to the DSS (conditions B and C) an exact McNemar's test determined that the difference in the proportion of low or high use at trial 1 and trial 2 was not statistically significant, p = 1.000.

Table 6.12: Percentage of high and low automated system use across trials 1 and 2

Trial 1		Trial 2	
Low use	High use	Low use	High use
44% (n = 11)	56% (n = 14)	48% (n = 12)	52% (n = 13)

With this data transformation, chi-squared test for association was conducted between conditions. A non-significant association between condition and use of DSS at trial 1, X(1)= .051, p = .821 and trial 2, X(1)= 1.989, p = .158 was found (see Table 6.12).

6.4.4 H3: Individuals will use DSS when task demands exceed cognitive capacity (when workload is high)

In total 25 participants had the option to use the DSS during the task. At trial 1, 18 opted to use the tool at least once during the task. For those who opted to use the tool, 10 (55.5%) reported high global workload, for those who did not opt to use the tool, 6 (85.7%) reported low global workload. At trial 2 this increased to 19 participants opting to use the tool at least once during the task. For those who opted to use the tool, 7 (36.8%) reported high global workload, for those who did not opt to use the tool, 1 (16.7%) reported low global workload. Global workload scores attained via NASA TLX were converted into either Low or High based on the median scores for the sample. This allowed a Fisher's exact test to explore if a relationship between use of DSS and workload was observed. The analysis found no significant relationship between these two variables at trial 1 ($p = .090$) or at trial 2 ($p = .073$).

H2b Global workload will be lower at trial 2 compared to trial 1

Workload was found to significantly decrease at trial 2 (5.07±1.85) compared to trial 1 (5.97±1.30); $t(37) = 3.681$, $p = .001$ (see Figure 6.6). This difference is possibly due to practice at the task decreasing the subjective experience of workload. Further, a Bayesian statistical analysis also revealed very strong evidence for the statistically important differences between groups (BF = 40.73), therefore we can accept the alternative hypothesis that there is a difference between groups (see Figure 6.7).

Figure 6.6: Bar chart depicting average global workload

Trial 1	1st Phase	2nd Phase	3rd Phase	4th Phase	5th Phase	6th Phase
A	6.60±1.08	6.32±1.03	6.20±0.99	6.24±0.88	6.07±1.07	6.16±1.02
B	4.95±1.42	5.27±1.43	5.31±1.55	5.54±1.40	5.73±1.72	5.81±1.88
C	6.17±1.11	6.16±1.55	6.19±1.71	6.25±1.61	6.35±1.38	6.22±1.59
All	5.90±1.38	5.91±1.39	5.89±1.46	6.01±1.33	6.04±1.40	6.06±1.51

Trial 2	1st Phase	2nd Phase	3rd Phase	4th Phase	5th Phase	6th Phase
A	5.65±1.74	5.72±1.90	5.62±1.87	5.63±1.76	5.47±1.77	5.44±1.91
B	4.01±1.83	4.17±1.97	4.21±2.17	4.03±2.09	4.27±2.22	4.26±1.88
C	5.61±1.54	5.68±1.69	5.77±1.43	5.44±1.61	5.34±1.79	5.21±1.86
All	5.08±1.84	5.18±1.95	5.19±1.94	5.02±1.93	5.02±1.97	4.96±1.90

Figure 6.7: Mean and SD Global NASA-TLX workload ratings across all conditions and each phase

Workload and total number of changed decisions

Independent t-tests were conducted to explore if a difference between the mean number of changed decisions and workload exists. At trial 1 no significant difference between means was observed $t(36)= .269$, p = .790. However, there were significant mean differences at trial 2, $t(25.812)= 2.318$, p = .029, Levene's test of homogeneity of variance was significant therefore equal variances not assumed is reported. On average participants who reported low workload changed their decisions more (M = 30.53 SD = 26.88) than participants who reported high workload (M = 14.68, SD = 12.84), 95% CI [1.787, 29.897].

Further independent t-tests conducted to explore if a difference between mean number of changed decisions to either correct or incorrect decisions at trial 2 were found to be not significant (p = .094). Table 6.13 provides the mean and standard deviations of changed decisions at trial 2 across reported low and high workload.

Table 6.13: Means and SD of the number of changed decisions to correct or incorrect at trial 2

	Decisions changed to correct		Decisions changed to incorrect	
	Mean	SD	Mean	SD
Low workload	11.89	11.19	2.95	3.47
High workload	6.74	6.78	1.47	1.38

Weightings of workload dimensions

Spearman's Rho correlations showed significant association between weightings given by participants at trial 1 and trial 2 across all subscales of their NASA-TLX responses: mental demand (rs= .530, p = .001), Physical demand (rs= .673, p = .000), Temporal demand (rs= .626, p = .000), Performance (rs= .612, p = .000), Effort (rs= .565, p = -.000) and Frustration (rs= .499, p = .001). This suggests that participants were experiencing the designed workload pressures (i.e. mental demand) similarly across both scenarios (see Table 6.14).

Table 6.14: Mean and SD of NASA-TLX weightings across both trials

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Trial 1	4.16 ± 0.86	0.55 ± 1.06	2.97 ± 1.42	2.74 ± 1.41	2.29 ± 1.06	2.26 ± 1.54
Trial 2	4.11 ± 1.13	0.45 ± 1.16	3.08 ± 1.28	2.71 ± 1.31	2.37 ± 1.15	2.32 ± 1.49

6.4.5 H4: Individuals in low accountability condition will select DSS (more times) compared to individuals in high accountability condition

A Kruskell Wallis analysis showed no significant relationship between the accountability condition of the participant or their subsequent accountability rating at the end of the task; in trial 1, $X^2(1)= 1.455$, p = .228 (mean rank scores Low = 17.22 and High = 21.55) and trial 2, $X^2(1)= .022$, p = .883 (mean rank scores Low = 19.75, and High = 19.22), suggesting that accountability was internally generated as opposed to being externally manipulated by the prime at the start of the experiment. This is confirmed by a significant relationship between self-reported accountability scores at trial 1 and trial 2 across all conditions, rs = .675, p = .000.

Due to all cell counts being less than 5, the number of times the DSS was selected was dichotomised into high use (as 2-3 times selected) and low use (as 0-1 times selected). Accountability was transformed using the median split (Mdn = 6.00 for both trials). This enabled a Chi-squared test for association to be conducted between accountability rating and use of DSS. There were no statistically significant associations between variables at trial 1, Fisher's Exact test p = .689 or trial 2 $X^2(2)= .987$, p = .320.

Accountability was also explored in its potential relationship to task performance. Pearson correlations showed no significant association between self-reported accountability scores and total number of correctly classified tracks at trial 1, r(38)= .300, p = .067 or trial 2, r(38) = .142, p =.396. Equivalent non-parametric confirmed no association between self-reported accountability scores and total number of correctly identified tracks at trial 1 (rs = .311, p = .0.58) or trial 2 (rs = .228, p = .168).

Pearson correlations showed a significant moderate negative correlation between self-reported accountability scores and total number of incorrectly classified tracks at trial 1, r(38)= -.480, p = .002, but no association at trial 2, r(38)= -.076, p = .649. Further a Bayesian statistical analysis showed very strong evidence that there was a negative correlation at trial 1 (BF = 35.69).

Independent t-tests found no significant differences between high or low accountability and total number of changes to decisions made at trial 1 (p = .285) or trial 2 (p = .765).

6.4.6 H5: Explore the relationship between scores on each cognitive trait and automation use

Automation use was dichotomised into early (selected in phases 1-3) or late (selected in phases 4-6) and each cognitive trait was dichotomised into high or low based on median split. Chi-squared tests were conducted however all cell counts were less than 5 due to the small sample size. Therefore, automation use was dichotomised into high (2-3 times selected) or low (0-1 times selected). This data transformation enabled chi-squared tests to be conducted to explore if an association between each cognitive trait score and automation use exists. None of the analyses were significant (p > .05) which suggests that scores on each of the six cognitive traits did not influence the decision to use the automated system or not.

6.4.7 H6: Explore the relationship between scores on each cognitive trait and task performance

Pearsons correlations were performed to assess if an association between the total number of correct decisions and each cognitive trait existed - all correlations were found to be not significant (p > .05). Focusing on the BIS-attentional subscale scores independent *t*-tests showed a weak mean difference between individuals who scored low on the subscale compared to high and the total number of correct decisions made at trial 1, *t*(36)= 1.974, p = .056. Individuals who scored low on average classified a total of 164 tracks correctly (SD = 64.19) compared to individuals who scored high on the scale who on average classified a total of 199.56 tracks correctly (SD = 74.58). However, this near significance was lost at trial 2 (p = .645).

Similar findings were observed when looking at the total number of incorrect decisions, except for scores on CF, and specifically CF-Alternative scale. At trial 1 CF was found to negatively correlate with incorrect decisions, r = -.419, p = .009, CF-Alternative, r = -.482, p = .002, however no significant association was found at trial 2 (p = .140).

Looking specifically at the identification of hostile tracks, Pearsons correlations revealed no significant associations between the cognitive traits and total number of correctly classified hostile tracks at trial 1. However, at trial 2 a negative correlation was found between NFC and correctly identified hostile tracks, r = -.515, p = .002.

6.4.8 Supplementary Data

Six themes emerged from the analysis of the questions relating to use of the system. Figure 6.8 display these themes and a selection of quotes from participant responses at trial 1. The most common theme to emerge was that of using the system to understand how it functions, followed by using the system to help deal with uncertainty relating to the decisions that needed to be made. Themes were related, for example;

I chose to use the decision support system when I became the most unsure about the classifications that I had made, also when I felt overwhelmed by the amount of unclassified icons on the screen. I chose to use it at these points due to being informed that it was as reliable as a well trained experienced operator, therefore I tried to use it as a method of backing up the decisions that I made. I also tried to use it at even spaced intervals throughout the task. I also sometimes, found it difficult to remember when it was necessary to classify an icon as neutral or friendly, so I decided to use the decision support tool when the uncertainty of this was highest. Toward the end of the task it became more overwhelming as I had classified most of the icons closest to the base. However, I then became unsure of the decisions I had made and started to doubt them, so I decided to use the decision support tool then.

This participant referred to the uncertainty felt around making classification decisions, opting to use the system to support their decision making and improve task performance. They selected the tool when they felt their workload and uncertainty increasing. This use of the system was expected; individuals could lean on the system when faced with high-levels of task-based uncertainty (endogenous uncertainty).

Although participants were informed that the system would not check their classifications a common rational written about selecting the system was to check the decisions they had already made. This suggests that participants did not fully understand the system functionality following a single training session. When introducing new systems to individuals it is vital that they comprehend the functionality so that the system will be used appropriately. The comments from participants suggest that a single, although clear, training session is not enough for full comprehension of a basic system.

System functionality

- “didn't require it to manage the task, but hadn’t used it and wanted to see what it did”
- “I thought I would be able to identify if it marked anything as hostile and I was going to use that to help me identify them in the future.”
- “Because I wanted to see what decisions the decision support system would make in regards to how it would classify each aircraft, to see if how I was classifying them was correct”
- “First Use: To determine how the system performed and to understand what position I would need to be in for it to be used. “

Uncertainty

- “One of the times I could not zoom in enough to be able to click on the unknown wing hiding behind a commercial fixed wing. The other time I used the decision support system was because I was uncertain.”
- “I chose to use the decision support system when I became the most unsure about the classifications that I had made”

Workload

- “There were too many airplanes to rate and I had reached a high level of mental effort, while I also hoped the decision support system would rate aircrafts I was not sure how to rate.”
- “Initially there was so much information on screen that it seemed overwhelming, therefore I used the automation system to try and reduce the quantity of information I needed to process.”
- “Second Use: The quantity of unknowns and the possibility of hostiles was causing a mental demand that was leading to stress/frustration. Third Use: The number of "need to check again before classifying" unknowns increased too far!”

Task Performance

- “It felt tactically to employ the decision support system at this point as I needed to reduce the amount of possible 'neutral' aircrafts so I could focus on any military aircrafts appeared”
- “I used it at the beginning to get me going so I could see what it analysed as friendly hostile or neutral. I could then look at these and make my own decisions on others.”
- “I choose to use it when I felt like I had classified ones that were in close proximity to the best of my ability. I thought that the system may pick up on any mistakes that I had made and correct them as well as identify the ones that I had not classified yet. This was why I mostly used it towards the end of the time limit, when I felt like the time was going to run out the system would do these ones for me. “

Time pressure

- “Time pressure made it obvious that I was going to struggle to classify all the targets, so I used it periodically to relieve some of that pressure.”
- “This was why I mostly used it towards the end of the time limit, when I felt like the time was going to run out the system would do these ones for me.”

Manual Preference

- “I felt it was my responsibility to allocate the aircrafts into the appropriate categorises and nobody else's. I also wanted to attempt the task on my own in order to see the behavioural patterns of the planes to make it easier for me to categorise them”
- “I didn’t use it because I felt I should complete it by myself.”
- “I felt it was my role to decide if situations were friendly, neutral or hostile - not the support systems”

Figure 6.8: Themes identified at trial 1

Figure 6.9 displays these themes and a selection of quotes from participant responses at trial 2. Although the same six themes emerged from analysis of the data the most common theme to emerge at trial 2 was using the system to help reduce workload followed by a preference to complete the task manually.

Workload remained a cited factor behind participants use of the decision support system even though average workload ratings reduced at trial 2 compared to trial 1. This highlights that workload remains a valuable performance metric to analyse when exploring the uptake of automated systems.

Fewer participants referred to using the system to "see what it does" due to having previous experience of the system from trial 1. However, several participants still referred to using the system to check their decisions, even though they were informed again in the training brief that the system does not check decisions already made. This suggests that a combination of training and experience with the system is required by individuals in order to fully understand how a system functions (Stanton and Ashleigh, 2000). Additionally, an increase in preference for manual control was cited by participants at trial 2. This supports the literature that posits that self-confidence in ones' own ability at the task will lower use of an automated system (Lee and Moray, 1992; Powell, 1991). Uncertainty remained a factor underpinning the participants' decision to use the automated system. Participants would complete as much of the task as they were comfortable with and defer to the system when facing uncertainty with their decisions, *"use the automated system in moments where I was more unsure about how to classify them"*.

Three themes emerged from participants responses to what they would do differently if they could perform the task again. Comments related to how they would use the system, their decision process or that they would not change anything as they were happy with their performance. There appeared to be a shift between trials with a greater reference to deferring to the system decisions that remained uncertain to the participants at trial 2. This corroborates the tendency towards manual preference voiced by participants at trial 2 as participants would only defer to the system if they felt they could not make the decision thereby avoiding making the decisions under uncertainty that they found difficult.

Additionally, participants were potentially making inferences about the system based on their ideas of how they would like and think the system should function, even after two training sessions and previous experience of using the system. For example, *"If there was a military craft and I couldn't see if it had originated from a friendly* *airport I would have used the decision system to make the decision".* This highlights that aspects of how the microworld was designed remained unknown to the participants supporting the opaqueness of ARCS and is also in line with the concept that individuals hold an automation schema which influences their expectations towards the system. It is suggested that a preconceived automation schema can influence what information is retained during training on system functionality which has important implications for training on the functionalities of automation systems.

System Functionality

- “I believe that automated systems are better at consistently getting these types of decision right than humans.”

Uncertainty

- “I chose to use the decision support system because I felt unsure on whether or not my answer was going to be correct.”
- “I feel I could/should have used this system when unsure about any hostile threats to ensure I wasn't classifying one on my side as a hostile”
- “There was one occasion where I most likely could have used it to help with the classification of a couple of tracks close together when I could not decide whether they were hostile or not. It may have been best to use the system rather than delay the decision and make a best guess.”

Workload

- “I got lost and thought it could help and it allowed me time to think of what was happening and get my mind back into it.”
- “Because of the mental demand, a big quantity of aircrafts to monitor and track and the hope the system would diminish the amount of aircrafts to rate and that it would rate them correctly when I was not sure of how to categorize some aircrafts.”
- “I used the system when I felt overwhelmed with the classification task. As the 5 minute segment breaks for questionnaires provided respite, I all too quickly pressed to carry on rather than pause and prepare for the continuation of the classification. The second (and third - if used, not sure!) use was when frustration levels increased above mental capability and the use of the DSS was to provide a 10 second respite.”

Task Performance

- “I do not feel that using the support system would have helped my performance”
- “I was struggling to assess whether some of the military aircrafts were 'friendly' or 'hostile' and so decided to employ the decision support system for some help and assistance at categorising these aircrafts”

Time Pressure

- I used the decision support system at intervals. I used it twice in the middle of the 30 minutes and once at the end. I did this because I knew I ran out of time last time so this time at the beginning I wanted to do it by myself to get a head start and then go on to use the support system.
- When I realised that it was working I employed it in the middle of the scenario to ease time pressure, and then in the last 30 seconds to catch any remaining unclassified tracks

Manual preference

- “There was not any time where I thought I should have used the support system because it is only as well trained as an experienced officer and I have now had practise at this: I didn't feel like I was unreliable.”
- “I did not feel I needed to use the decision support system an additional time as I felt confident in my own ability and performance during the task”
- “I wanted to try and have a go myself and sometimes support systems if they aren't told what they actually do could do more harm than good.”
- “I wanted to classify most of the icons myself and then see if they were correct using the decision tool”
- “I did not opt to use the decision support system simply because I preferred to classify myself to see how accurate I was”

Figure 6.9: Themes identified at trial 2

6.5 Discussion

This chapter presents the six hypotheses relating to possible motivational and cognitive factors behind using an automated decision support system (see Figure 6.6). The analysis of the data raises interesting questions that will now be discussed.

6.5.1 H1: Individuals with access to DSS during task (Groups B & C) will perform better (i.e. correctly classify more tracks- classify all hostile tracks) compared to control group (A)

In opposition to H1, it was found that performance did not improve significantly for participants with access to the DSS. These findings support the smaller body of literature that posits that automated systems do not always mean task performance improvements (Beck et al., 2007). As there were no conclusive benefits of using the system these findings have important implications for how future research should approach exploring the use of automated decision support systems. As Osiurak et al. found (Navarro and Osiurak, 2015; Osiurak et al., 2013) individuals prefer manual control if the system cannot improve upon task speed. A similar rhetoric was revealed from participants at trial 2 with a greater number of comments relating to their disuse of the system due to feeling confident in their own ability to complete the task. These findings support the literature that has shown how individuals compare their abilities to that of the system when making a judgement on whether to use an automated system or not (Powell, 1991). Therefore, these findings highlight the importance of making the value of the system transparent to the operators.

A significant improvement in task performance from trial 1 to trial 2 (on the neutral and total classifications correctly made) was found. This could be a result of participants prior experience of the task leading them to assume similarities between the scenarios in the number of expected neutral tracks they would observe, thereby classifying more neutral tracks the second time they completed the task. However, improvement in correctly identifying more friendly tracks was only observed in the control and high automation group and across all conditions there was not a significant improvement in the number of correctly identified hostile classifications. This is in line with literature that shows how individuals can improve in their decision making when dealing with expected observations but do not improve in performance when making decisions on unexpected observations (Baron, 1993). That is to say, participants were expecting a high number of neutral tracks when they completed the task a second time as they had experienced a high number of neutral tracks previously. This is supported by the analysis of the number of neutral tracks classified each phase where at trial 2 on average more neutral tracks were classified in the first 5 minutes of the experiment compared to the later stages of the scenario. However, as the number of hostile and friendly tracks made up 5% and 2-3% respectively of each scenario, these types of tracks were not expected - mirroring reality. Therefore, although friendly track identification did improve in two groups, possibly due to participants in those group paying closer attention to aircraft that appeared around the designated "friendly airbase", significant performance improvements were not observed across the board, and not at all with identifying hostile aircraft.

Anecdotal evidence was found for the benefit of using the decision support system on the total number of correct hostile classifications at trial 2 when looking specifically at the participants who had the choice to use the system or not. It is therefore tentatively suggested that if provided with the option to utilise an automated system to aid performance, those who opted to use the system were able to correctly identify more critical (i.e. hostile) tracks compared to those who did not utilise the system. As the system would only classify neutral tracks, utilising the system may have allowed participants time to recheck previously made classifications or to scan the environment and therefore see potential hostile tracks appear. When they took back control they were then able to select and correctly classify those tracks.

This study utilised a unique design in that participants were presented with a single realistic (i.e. high complexity, low probability of criticality and inherently uncertain) scenario and were given the option to decide to use the automated system or not. Therefore, the finding that task performance was found to improve for the high probability tracks but not for the low probability tracks and that use of the system, in some cases, supported the participant to identify a greater number of low probability tracks has implications for operational environments. It is the low probability events that carry the greatest risk to personnel in operational environments and therefore the quick identification of these events is crucial. These findings support the literature that shows that individuals develop biases to high probability events which can result in cues to low probability events being missed. It should also be noted that these results could be an artefact of participants being novices at the task. Although all participants completed training in the task there was limited exposure to low probability events during the training, with more experience of low probability events detection of these events during the task may improve.

6.5.2 H2: Individuals in low automation condition will select DSS more than individuals in the high automation conditions

H4 was not supported by the data as no significant difference was found between the low automation and high automation conditions and selection of the DSS. However, the rationales provided by participants as to why they used the automated system suggest that the information provided about the system was a factor behind these decisions. For example, *"I chose to use it at these points due to being informed that it was as reliable as a well trained experienced operator, therefore I tried to use it as a method of backing up the decisions that I made"*, this quote highlights that this participant, from the low automation condition, was using the system as a team-mate, to corroborate their decisions. Further one participant from the high automation condition stated, *"I believe that automated systems are better at consistently getting these types of decision[s] right than humans"* they were also rated as a high user of the system. Although anecdotal, these findings support the notion that previous understanding and expectations of systems do influence later decisions around system use and therefore it is vital that future research focuses on how to measure these expectations and to match system functionality to them in order to support appropriate system use. The findings from this study are in line with previous literature which has posited the influence of previously held views towards automation and uptake of an automated system. Further supporting the importance of training with a system prior to use. Detailed training and the transparency of the system will enable the operator to develop a full awareness of the value the system can provide to task completion; this knowledge will then become a motivational factor behind uptake in operational environments. Additionally, there were no differences in use across the two trials, suggesting that participants will continue to use an automated system in a similar way each time they complete the task. This finding highlights the importance of facilitating accurate understanding of functionality and effective implementation of a system during its first application, as future use of the system will be closely related to this.

6.5.3 H3: Individuals will use DSS when task demands exceed cognitive capacity (when workload is high)

As expected workload was found to significantly decrease with practice at the task. However, contrary to H2 workload was not associated with decisions to use the automated system. Although no statistical support was found for H2, the rationales provided by participants following the main task suggest that perceived workload was a factor in the decision to use the automated tool or not. Additionally, over 50% of the participants who chose to use the automated system during trial 1 reported high workload on average across the task. Therefore, this study supports the previous literature that has discussed the value of exploring workload and its impact upon automation uptake and use.

It was also found that during trial 2, participants who were experiencing low global workload were able to recheck more classifications, thereby making more changes to their decisions compared to the participants who were experiencing high global workload. No differences were found in relation to the total number of decisions changed to correct decisions or incorrect decisions. Low workload allowed greater verification behaviour via checking previously made decisions, although the analysis did not clearly show that increased checking led to an increase in correct decisions. Providing operators with systems or training which enables them to deal with task demands and lower workload is one possible avenue which could mitigate against the pitfalls of biased decision making towards high probability events over low probability ones.

Global workload remained relatively stable across each phase and the desired workload metric of mental demand was produced by both scenarios. Additionally, both scenarios were deemed to be of similar difficulty, which replicates the reality of operating on a ship in the same location for extended lengths of time. Although the initial design plan was to cause an increase in workload during the critical incident, it was decided to instead maintain workload levels as would be expected in the real world, where the criticality of the environment can gradually increase. It would be of interest to measure global workload of personnel daily when they are at sea to confirm, if as posited by this study, that workload would remain relatively stable. Therefore, the fact that workload was not found to be associated with using the automated system supports the argument of Schwarz et al. (2014) that research should look beyond just using workload as a user assessment measure.

6.5.4 H4: Individuals in low accountability condition will select DSS (more times) compared to individuals in high accountability condition

Accountability was found to be an internally generated construct as opposed to being externally manipulated by the primes given to participants, supporting the fact that accountability can be a challenging variable to externally manipulate (Skitka et al., 2000). From this finding it is posited that the concept of felt accountability which refers to an individual's perceptions of their own accountability (Frink and Klimosky, 1998) was therefore explored in relation to decisions to select the automated system or not. No significant relationship between accountability and selection of DSS was found, therefore H3 is not supported by this data.

A significant association between self-reported accountability and incorrect decisions was found at trial 1, participants who felt greater accountability made fewer errors at the task. This is in line with research that posits that individuals who feel greater accountability will be more careful when making decisions (Beck et al., 2007). However, as this association was not found at trial 2 and high levels of accountability were also not associated to increased accuracy of decisions, as would be expected with increased carefulness when making decisions, this finding should be interpreted with caution.

A recent review on accountability literature has highlighted the potential influence of social desirability on self-reported ratings of accountability (Hall et al., 2017). Participants in this study were asked to rate their feelings of accountability on a scale of 1 to 10 and provided a range of ratings suggesting that this question was not necessarily influenced by social desirability but that participants were providing an honest answer. The fact that accountability ratings across the two trials were correlated within participants further suggests that participants provided an honest rating of how accountable they felt. However, in operational environments where there is a hierarchical structure accountability may have a more overt influence on automation use or disuse. For instance, literature has linked the importance of understanding one's own role in a team so that their accountability can be adjusted to that role (Adams and Webb, 2002; Onnasch, 2015). Due to the difficulties of manipulating accountability experimentally it would be of interest for future research to explore the potential influence accountability has upon automation use or disuse in the field. It would also be of interest to develop ARCS to increase external pressures and thereby generate greater levels of accountability. This will allow the robust testing of new systems and how they may influence accountability and equally how accountability may influence their use or disuse in a safe environment (Alison et al., 2013).

6.5.5 H5: Explore the relationship between scores on each cognitive trait and automation use

Strong associations were found within participants for their scoring of each cognitive trait measure which suggests that these measures were assessing trait characteristics. However, there was found to be no association between scores on the six cognitive trait measures and automation use.

6.5.6 H6: Explore the relationship between scores on each cognitive trait and task performance

Task performance was not found to be reliably associated to scores on each cognitive trait. That is, no relationships were observed across both trials.

The results from this experiment tentatively posit the influence of attentional impulsivity on degrading task performance. The results from trial 1 suggest that individuals who score highly on this subscale of the BIS-11 performed poorly in comparison to the individuals who scored low on this subscale. Supporting research that has found an association with impulsivity and attentional lapses (Levine et al., 2007). However, the same difference was not observed at trial 2, it is posited this is due to the dramatic increase in performance across all individuals at trial 2 overshadowing the interaction between attentional impulsivity and task performance. It could be that experience at the task mitigated the detrimental effect attentional impulsiveness can have upon task performance, highlighting the importance and value of experience and training.

It was also found that individuals who scored highly on CF, and specifically the CF-Alternative scale made fewer incorrect decisions during trial 1. However, this association was not observed at trial 2 therefore these findings tentatively support the literature that has shown how individuals who score highly on CF and CF-Alternative scale are able to make fewer incorrect decisions under uncertainty and therefore are able to adapt to uncertain environments.

Individuals high on NFC were found to make fewer correct hostile classifications at trial 2, this may be due to experience of the low frequency of hostile tracks at trial 1 influencing expectations of the task and therefore resulting in information that supports the presence of neutral tracks being seized upon early. This resulted in alternative evidence or later information being ignored as individuals had already decided on track classification.

6.5.7 Implications of findings

The findings from this study were unexpected, particularly that task performance was not improved with access to the automated system. There are several considerations that may account for why the findings from this study oppose the vast majority of the literature that has explored human-machine-interaction. Firstly, the majority of experiments exploring human interaction with automated systems compress the 'significant' events into short scenarios (for example see Beck et al. (2007); Dzindolet et al. (2003); Helldin (2014)). However, in real life personnel may go hours or days without a significant threat and therefore creating experiments that elicit a 'snapshot' of the overall task do not provide a complete picture (Hutchins, 1997). Therefore, the findings from this study provide a novel and more realistic understanding of how people interact with automated systems when conducting a task that takes time, has elements of uncertainty and criticality but also is a task that, with practice, individuals in general feel capable, and are able to complete.

Although under time pressure, for instance participants had 30 minutes to classify over 200 tracks, they were not subjected to more overt time pressure i.e. you must make x number of decisions in y minutes. It may be of interest to run ARCS using a scenario where the time pressure on each decision is more overt to participants to see if interaction with and rational for using or disusing the system changes. However, if the complete reality of the operational environment is not considered by researchers then the findings from these studies will only be applicable in the limited instances were operators are having to make decisions under extremely constrained parameters. The wider literature on automation use posits that trust in the system is paramount to enabling appropriate system use, however the findings from this study suggest that during daily operations personnel will be capable of managing the classification task manually. The potential implications of this are that when the operator is then faced with an impossible task (due to time pressures or the number of incoming threats) they may opt for the automated system not out of trust but out of a lack of choice. Therefore, this brings into question whether current literature is providing findings that are helping to support human-automation interaction across all circumstances or only during situations when the choice of manual control is taken away from the human.

The use of a student sample for this study should also be considered when interpreting the findings. All students were novices at the task prior to taking part but as was seen with the data analysis, all participants displayed learning effects the second time they completed the task. Within the Royal Navy (RN) lower ranking personnel will complete this classification task to build the recognised air picture. The majority of these individuals may not have been to university as they will have joined the RN out of school or college. Therefore, arguably university students should have the mental aptitude to be able to learn how to perform, a simplified version of this task effectively, as was evidenced by this studies sample. However, it would be of interest to run this study with a sample of RN SME to explore how they interact with the automated system.

Out of the total number of participants who had access to the system 72% (n= 18) opted to use the system in trial 1, and 76% (n=19) opted to use the system in trial 2. The high percentages of participants choosing to use the system is contradictory to previous literature that highlights a general preference for manual control (if it is an option) over letting an automated system complete the task (Beck et al., 2007; Lee and Moray, 1992). For example, Navarro and Osiurak (2015) found that for a preference towards the automatic completion of the task to be found, the tool had to operate at a perceived 4x speed increase. However, the different experimental design employed by this study may explain these findings. The system would only run for a short duration (10 seconds) therefore participants may have perceived that the tool would not take over the task thereby allowing them to maintain control of the task and still use the system. Alternatively, the high rate of system use could have been an artefact of participants not understanding how the system functioned, thereby selecting the tool to see what it does. This fact is referred to by several participants in their feedback following the task.

Themes identified from supplementary data should be taken and developed further. For instance, workload was reported by participants to be a factor behind their decision to select the automated system or not regardless of the level of workload reported. Which suggests that not all motivational influences can be easily measured and analysed quantitatively highlighting the value of collecting qualitative rationales alongside quantitative metrics when exploring the uptake of automated systems.

The common theme of using the system to see what it does highlights the importance of involving the individuals who will be using the automated system in the development and designing stages, as well as highlighting the importance of training. Koltai et al. (2014) posit that appropriate calibration to an automated system can be obtained through exposure and experience of real-world scenarios. The findings from this study support this proposition. To improve skill an individual must be exposed to training in high-validity environments and have adequate time and opportunity to practice at the task (Kahneman and Klein, 2009). This is also true for developing an understanding of how automated systems function and highlights the requirement for the development of high-fidelity test beds that will allow operators to learn the new skill of completing the task alongside an automated system. A lack of skill can result in incorrect intuitions as was seen with participants. Although participants were given clear instructions on how the system operated, for example that it would not check their classifications but would make new classifications, participants incorrectly tried to use the system to check their decisions. This suggests that how new systems are introduced to operators and how training incorporates them into operations is crucial to prevent misunderstanding on the role of the system. These findings could however be an artefact of using a simulated environment as it has been found that participants may focus on information searching over action implementation (Alison et al., 2015). That is, participants did not just use the information presented in the training brief but felt the need to use the system once to see for themselves what it did. Additionally, it may be that participants were expecting the system to function in a particular way, through holding an "automation schema" (Dzindolet et al., 2002; Rice and McCarley, 2011) therefore they were cognitively biased to expecting the system to check their decisions as opposed to hearing how the system actually functions. A relationship has been found between expectations in a system pre-use and reported experience with the system post-use (Li et al., 2008). With research arguing for human centred design to enable systems to be built in accordance with operator expectations (Ashleigh and Nandhakumar, 2007) future research should continue to explore the influence expectations have upon interaction with automated systems.

6.5.8 Limitations

Participants remained unaware of how ARCS functioned which confirms the opaqueness of the microworld. The scenarios presented to participants ran for 30 minutes and were complex in their design supported by participants reporting a consistent level of workload during the experiment. Finally, ARCS ran in real time therefore fulfilling the three characteristics identified that constitute a valid microworld (Alison et al., 2013; Brehmer and Dörner, 1993; Chen and Bell, 2016). Through the iterative development of ARCS using the knowledge gained from conducting the study presented in chapter 3 and through informal conversations with experts, ecological validity was built into the interface and the scenarios. However, there is divergence from the reality of the task that ARCS was simulating as the task is usually conducted by a team of individuals working on the bridge for at least 4 hours. It was deemed unrealistic to ask university students to spend over 4 hours in an experiment therefore the ARCS experiment designed for this thesis ran for 2 hours. However, this length of time and the repeated measures nature of the experiment negatively impacted upon the ability to recruit participants. Although incentives were available, for example course credit or payment, these incentives were not enough to ensure the aimed for sample size was reached. However, the findings have important implications due to the chosen study design. Participants were required to complete a task that was complex and long as opposed to the widespread practice in human-machine-interaction literature when exploring threat detection of using blocks of short scenarios where arguably participants are artificially primed to take action. Therefore, the findings are more generalisable to tasks that are currently undertaken by personnel. Additionally, self-reported questionnaires rely on participants possessing awareness of themselves, i.e. insight into their personality (Bari et al., 2016) and can be vulnerable to social desirability biases. However, the fact that scoring by participants on each cognitive trait scale was reliable across both trials suggests that social desirability biases were not observed in this study.

The scenario participants completed is specific to the first stage of conducting the air defence task, the development of the recognised air picture. Therefore, the findings from this study may not transfer to other tasks that are conducted on-board naval vessels, for example using an automated system to aid in navigating the vessel. However, lessons can be passed on to similar tasks that involve making decisions under uncertainty when dealing with high cognitive load and managing latent time pressures.

6.5.9 Chapter Summary

This chapter explored how several factors, motivational and cognitive, interact with the decision to use an automated system or not and task performance. A high-fidelity microworld experiment was run on a population of university students with the findings highlighting that the use of automated systems is not always associated to task performance improvements. Training effects were observed with task performance and participants were found to utilise the automated decision support system in an equivalent manner across both trials. These findings have interesting implications for the literature that has argued the importance of transparency of system functionality as the results presented suggest that individuals will approach uptake of an automated system in a consistent way. Therefore, if inappropriate uptake occurs initially it is likely that this inappropriate use will persist. The findings from this study also highlight the necessity of taking a mixed-methods approach to this field of research. For example, although workload was not statistically found to influence uptake of the decision support system it was commonly cited as a factor that impacted upon participants decisions to select the tool as evidenced by the supplementary data.

Chapter 7

General Discussion

7.1 Introduction

The aim of this thesis was to provide a novel contribution to the literature on human-machine-interaction in the context of military operational environments. Specifically, this thesis adopted a mixed-methods interdisciplinary approach to: (i) present the decision-making stages of the air defence task; (ii) discuss how automated system are currently used in operational settings and where they may be used in the future; (iii) discuss the existing procurement process, highlighting the barriers to effective automation application; and (iv) develop a high-fidelity microworld to explore individuals' rationales of using a generic automated system when performing a threat detection task.

Qualitative methods were adopted to holistically explore the decision stages of the air defence task, presented in Chapter 3. The same methods were again used to gain contextual depth and breadth of understanding on the current operational use of automated systems by Royal Navy (RN) personnel and where personnel view the use of automated systems will have the most benefit in the future. Additionally, and uniquely, personnel were asked for their thoughts and experiences of the current procurement process, Chapter 4 presents this study in full. Identified through this qualitative research was the importance of developing and using simulated environments to explore how individuals interact with automated systems. This was achieved using quantitative methods with the development of the Automatic Radar Classification Simulation (ARCS) and the running of a within subjects student experiment, presented in Chapters 5 and 6. The central contribution of this thesis was to: (i) update the understanding of the decision stages of the air defence task in light of increasingly sophisticated weapons technologies; (ii) identify the negative impact of lack of communication and collaboration between personnel and system designers on automation uptake and use; and (iii) to develop a micro-world of the initial stages of the air defence task that can provide a robust simulated environment to test the interaction of human operators with automated decision support systems. Therefore, this concluding chapter will:

(i) Provide an overview of thesis chapters

(ii) Discuss the main findings of this thesis and their implications

(iii) Identify and discuss the methodological strengths and weaknesses alongside recommendations for future research

7.2 Overview of thesis chapters

- Chapter 1: outlined the psychological underpinnings of this thesis drawing on the Naturalistic Decision Making (NDM) and heuristics & biases literature to empirically ground the research conducted for this project. Additionally, the chapter provides an overview of the literature into human-machine-interaction to provide context to the problem space under exploration

- Chapter 2: presents the mixed-method framework developed and followed by this thesis and highlights the challenges typical to interdisciplinary research, such as access to Subject Matter Experts (SME).

- Chapter 3: presents the first qualitative study. RN SME completed a questionnaire that explored the decision making stages of the air defence task. The findings highlight that the high-level stages of the task; Observe, Identify and Classify, and Decide and Act, have remained the same over the last 30 years.

- Chapter 4: presents the second qualitative study which explores current automated systems and how they are used operationally by RN personnel. Additionally, this chapter presents where personnel view automated systems will play a role in the future as well as opinions towards the procurement process. It was identified that often systems are released into service not fit for purpose. This has a detrimental impact on capability and increases the financial costs associated with bringing systems into service.

- Chapter 5: describes how ARCS was initially tested with two pilot studies. These pilot studies confirm the ability of ARCS to provide research with a high-fidelity microworld to explore automation use decisions in the maritime environment.

- Chapter 6: presents the findings from the student experiment run with ARCS. The findings of this study show that the introduction of an automated decision support system does not necessarily lead to task performance improvements. Additionally, that the mixed-methodology adopted enabled a holistic understanding of *when* and *why* individuals were selecting to use an automated decision support system.

7.3 Summary of main findings

7.3.1 Operational use of automated systems in the RN

The central contribution this thesis makes is in exploring the operational use of automated systems in the RN. Revisiting the air defence task in Chapter 3 to update the descriptive model of this task, it was revealed that the key stages of this complex duty have remained the same over the last 30 years. Additionally, the identical areas identified by Holt in 1988 that could benefit from the application of automation remain of importance for current RN personnel. For example, Force Threat Evaluation and Weapon Allocation (TEWA) remains crucial considering the continued requirement to work across nations. These findings raise pertinent questions as to why the same areas are still in need of support. Therefore, the next step of this thesis was to explore the operational use of automated systems by personnel today - what systems are available to personnel and how do they impact upon operations. Chapter 4 presented detailed qualitative findings that highlighted the ubiquitous nature of automated systems in tasks performed by RN personnel across the domains of above water, underwater, mine countermeasures and land and littoral manoeuvres. However, a critical finding of this thesis was the disconnect that exists between system designers and the personnel who will be required to use the systems developed. As discussed in Chapter 5, academia is uniquely placed to provide open architecture environments that are designed to promote collaboration and knowledge sharing between system designers and end users. Academic institutions are less constrained by commercial competition compared to industry partners and are organisations that foster a culture of collaboration and interdisciplinary research. This thesis has evidenced that this culture is required to understand how automated systems currently interact and, in the future, will interact with human decision making.

Highlighted just over 30 years ago by Tolcott and Holt (1987) the development of new automated systems tends to be driven by popular technologies as opposed to the systematic and holistic evaluation of the user's needs. This results in systems that are not always necessary, are overly complicated or are not fit for purpose. This thesis has evidenced that this is still often the case which can have detrimental effects on capabilities as well as increase the financial costs associated with procuring new systems. With the current austerity policy that the UK has been operating under since 2010, the fact that the current procurement process can result in increasing the financial costs associated with the development of new automated systems to support capability is concerning. In order to enable continued development of new systems to support military capability it is essential that the procurement process is improved to deliver systems that are both cost and performance effective. These factors also contribute to incorrect use and/or disuse of current and new systems. However, it is worth remembering that not all decision avoidance is irrational (Anderson, 2003), if the systems do not perform a required function it is understandable why they will not be used. Similarly, if automated systems are overly complicated and opaque, personnel may not fully understand the system functionality and therefore when and why they should be utilising the system. Thus, the findings of this thesis support the argument for designing transparency into automated systems to facilitate operator's understanding of functionality (Colebank, 2008; Johansson, 2010). The findings also highlight the vital requirement for developers to work collaboratively with the end users *throughout* the design, development and implementation stages to ensure that new systems are fit for purpose and utilised appropriately.

Literature has shown that effective communication within multi-agency systems is required to support situation awareness and decision-making (Waring et al., 2018). A further key finding of this thesis is the importance of effective communication. Chapter 3 supports this by highlighting the need for effective communication within the operations room team to facilitate accurate development of the recognised air picture. Ineffective communication can result in increasing uncertainty around situational awareness. As all threat evaluation decisions are based on Situation Awareness (SA), if the SA is uncertain or incorrect this will negatively affect accurate decision making. Examples provided in this thesis highlight the consequences inaccurate decision making can have in this environment, for example the USS Stark incident. Additionally, it is accepted that communication is vital to developing shared understanding and common frames of reference that are required in order for the team to function effectively. Chapter 4 highlighted the requirement to facilitate communication between system developers and end users to support the appropriate development of new automated systems. To achieve this, innovative approaches must be developed and evaluated, for example the Athena project highlighted in Chapter 4. Currently these channels do not appear to be robust and therefore system designers and end users are not developing a shared understanding which results in erroneous system development and use.

Uncertainty was a critical factor identified by Chapter 3 as impacting on the decision-making process within the air defence task, supporting the wealth of literature that highlights the ubiquiousness of uncertainty in military environments (discussed in detail in Chapter 1). Chapter 3 also identified the importance of clear communication to mitigate against continued uncertainty in environments that are characterised by time pressures and criticality. This finding supports the literature that has highlighted the importance of clear and effective communication in multi-team system functioning in critical environments (Waring et al., 2018). Evidenced by Chapter 4 is that uncertainty can be increased with the introduction of new automated systems that do not function as expected or required which can lead to disuse. Yet, in Chapter 6 uncertainty was also found to motivate automation use in order to avoid dealing with uncertainty. The automated decision support system facilitated the omission of decisions that were perceived to be highly uncertain. These findings suggest that uncertainty plays a fundamental role in influencing decision making and use of automated systems. This role is context dependent and complex therefore future research must continue to take into account the effect of uncertainty to mitigate against problems it can cause (i.e. decision inertia).

7.3.2 Uptake of automation

Operator error has been commonly cited as the primary cause of sociotechnical system failure (Driskell and Salas, 1991) and has been associated with how complex the system is. For instance, the more complex and opaque the system is the increased likelihood that automation misuse or disue will occur. The findings from Chapter 6, where it was observed that participants opted to use the decision support system to see how it functioned, further support the importance of system transparency in facilitating automation uptake. However, it is not just transparency with regards to system operation but also in terms of doctrine and guidelines of system use that is crucial to supporting appropriate uptake and preventing system failure. Continued attention is required by organisations to develop clear guidance on the ethical stance towards incorporating automated systems into daily use. This guidance will feed into how accountable operators perceive their actions to be which will influence automation uptake. Additionally, systems must operate in a transparent way so that if a fault occurs or an error is made it is possible to understand what went wrong or why the system opted for the decision it did, this will only be possible if systems can be understood transparently. Academia can support the development of this guidance through continuing to explore organisational cultures and how these cultures are changing. This thesis suggests that it is vital to continue to explore sociotechnical organisations as technologies continue to develop and be introduced. Understanding the organisational culture and the influence this can have upon individual and team decision making is required in order to identify the facilitators and barriers to effective decision making in complex environments.

Additionally, identified by the findings of this thesis is the importance of facilitating adaptability to make effective decisions. Team decision making requires the actor to make assumptions about what the other members are doing based on what the others know, believe or want - the actor must take the perspective of the other team members (Boland et al., 1992). This ability to anticipate what other team members may need is crucial for superior team performance and adaptability to task demands (Serfaty et al., 1998). If the team consists of human operators and an automated system that functions as a 'black-box' and not in a transparent way it is difficult for the operator to understand system functionality and therefore make assumptions in order to supervise the system's actions properly. Adaptability can be supported through experience and training therefore the importance of facilitating immersive and frequent training opportunities for all actors in the sociotechnical system is once again highlighted. A central contribution of this thesis is the development of ARCS as a flexible and valid microworld that future research can utilise to explore human-machine-interaction in the maritime domain.

One main finding from the ARCS experiment was that task performance, and critically identification of hostile tracks, was not improved with access to the decision support system. This highlights that the introduction of automated systems may not necessarily lead to immediate task performance improvements, as was also reported by RN personnel in Chapter 4. This finding further highlights the need to ensure that systems that are developed will provide end users the support needed to increase capability. Just because it is now possible to automate most tasks, does not mean that all tasks need to be automated. It remains of interest to explore and develop robust training programmes that support the development of operator's adaptive expertise which enables them to notice the non-routine (i.e. increasing their own performance) as well as to function effectively within a sociotechnical team (i.e. the team which consists of the operator and the automated decision support system). Research has shown that specification of learning objectives are required at individual and team levels (Salas and Cannon-Bowers, 1997) and therefore training programmes must take into consideration how these objectives interact, as well as making sure each objective is addressed during training. This will enable the sociotechnical team to acquire skills, knowledge and a sense of efficacy, all of which are argued to support adaptability (Chen et al., 2005). Additionally, as the common method of training automated systems is by using vast data banks of "known" patterns of life and information, how system developers train awareness of the "unknowns" into systems remains of interest. For instance given that the last time the RN fought a comprehensive air battle was during the Falklands War (2nd April 1982-14th June 1982) there is a paucity of data that contains current potential "unknown" threats. Training how to extrapolate the "unknowns" into potential knowns will enable human operators and automated systems to remain robust and effective when facing contradictory and novel stimuli.

It is interesting that the results from the student experiment were unexpected and do not support the common arguments that come from the literature on human-machine-interaction. It is posited that the findings stem from the ecological validity of the task. As already discussed the scenarios presented did not induce fluctuations in workload to effectively mimic the reality of the operational environment on a normal day. Additionally, participants were explicitly asked for their rationales behind using the automated system which revealed the influence of commonly mentioned factors in the literature such as uncertainty and workload behind their decision to utilise the automated system, even though these factors were not clearly identified by the quantitative analysis. In contradiction to previous research this thesis found that levels of system uptake were high among participants and that uptake was approached similarly between trails. It is posited that the high levels of uptake were due to participants maintaining a sense of control over the task as it was their choice to use the system or not, a system was not imposed on their decision-making process. Therefore, participants were able to utilise the system to support and complement their decision making as opposed to being asked to use a system that takes over a certain stage of the decision-making task.

7.4 Implications and recommendations

Identified by the literature as a flaw of research on traditional rational models of decision making is the failure to take into account contextual factors (Eyre et al., 2008a). A similar flaw can be attributed to the traditional literature on human-machine-interaction. The stripping away of contextual factors to allow clear distinctions to be made of task levels has removed the complexity and intuition that influences how humans make decisions and interact with automated systems (Jamieson and Skraaning, 2017). This thesis has evidenced through qualitative research (discussed in Chapter 4) that traditional research into and design of automated decision support systems does not always support better 'in the field' decision making, as new systems are not designed holistically and therefore are not fit for purpose. There is however a new movement away from the traditional Level of Automation (LOA) and abstraction of task components. The findings from this thesis support the need for this new approach to understand how automation interacts with human decision making. The role of the human remains critical to ensuring that systems are utilised effectively, humanely and ethically, thus they cannot be an afterthought of system designers. Developing mental models and management of attention are critical skills to learn to function effectively in a sociotechnical environment therefore more attention should be given to including the end-users in the design and development stages of new automated systems. Chapter 4 also evidenced that there is a desire from the end-users of systems to be included in their development.

There is great value in research from context-rich environments that can produce credible and transferable conclusions (Mishler, 1990), therefore attention must also be paid to supporting more pragmatic and holistic research into human-machine-interaction. Consequently, this thesis adopted an NDM approach to the research which enabled the extrapolation of real-world recommendations from the findings that are presented. The collection of data from two qualitative studies with a unique body of experts has opened an important window into understanding the operational use of automated systems by the RN. Concern has been raised with only listening to experts when exploring organisational practice and decision making. By only focusing on a limited group of experts this can produce an incomplete picture of organisational practice (Gore et al., 2006). Cognisant of this, a large body of RN personnel were invited to take part in stage 2 of this thesis representing a range of experiences, job roles and levels of expertise ($N = 53$). Therefore, a detailed and novel picture of organisational practice across various levels of the RN was drawn from the analysis presented in Chapter 4. It is typical of industry to use "in-house" experts to provide recommendations and expert knowledge when undertaking research or developing new technology. Whilst these individuals do possess expertise, a key recommendation of this thesis is to use multiple sources of expertise. That is, to collaborate with multiple experts who have *current* experience of organisational operations at multiple hierarchical levels. More extensive collaboration is required to support the development and more importantly the effective integration of automated systems into operational use.

This thesis has drawn from the literature into automation and decision making. The findings have revealed that although the NDM community has developed and advocated for approaches that support decision centered design there remains a disconnect between system developers and the end users (evidenced in Chapter 4). Often there is also the inability from the literature into automation to state what the individual can achieve without an automated option within the study design. This results in a skew towards perceiving the application of decision support systems to any problem as beneficial, which has arguably led the research community to focus primarily on how individuals can be made to 'trust' automation. However, as evidenced by this thesis in Chapter 6, individuals are capable of performing complex tasks and were found to not significantly benefit from having the ability to use an automated decision support system. This thesis therefore recommends that the first step to any automated decision support system development is not to assume that the individuals require automation to perform their task better but should be to explore if the individual would actually benefit from being given the option of an automated decision support system.

Research has shown that incorrect mental models can also hinder learning (Rouse and Morris, 1986) therefore the development of high-fidelity safe to fail environments within which personnel can practice the task whilst cooperating with the new systems is vital. This thesis does not suggest that it is the first to advocate for the concept of utilising high-fidelity training environments. However, such environments can often be overlooked due to the expense of creating them and the time it takes to develop and run effective training scenarios which enhance transfer and adaptive capabilities (Kozlowski, 1998). However, such environments will provide users with a sense of typicality which functions as a baseline. This will enable abnormalities to be identified (Cannon-Bowers and Bell, 1997; Klein et al., 2005) thereby facilitating system error detection. Additionally, this thesis has highlighted the prevalence of Doyle's catch, that is that developers are creating automation that works for one situation within idealised settings but when that automation is implemented into an operational environment it lacks robustness (Woods, 2016). Therefore, not only would the increased application of immersive high-fidelity environments support operators' training they would also support the robust development of systems that are not brittle when used in the real world.

As previously mentioned, a further central contribution of this thesis to the wider academic domain is the development of ARCS. Due to the flexibility that has been designed into ARCS this microworld can be used to explore research questions that lie beyond the scope of this thesis. For example, future research could develop scenarios that involve team structures and decision making to explore the uptake of automation across teams. Additionally as ARCS can handle bolted on automated systems, future research can look to utilise ARCS to test newly developed systems in a robust way.

7.5 Methodological strengths and weaknesses

The next section of this discussion chapter will present the methodological strengths and weaknesses of the research approach taken. The majority of the data collected for this thesis was qualitative in nature and an inductive approach to analysis was taken. This resulted in the conclusions drawn being strongly grounded in the data, a strength of this thesis. Additionally, in order to mitigate concerns that qualitative research is less scientific, where possible inter-rater reliability was conducted to support the reliability of the analysis and therefore the findings. Presented in Chapter 2 is a clear overview of the mixed-methods approach taken by this work. Providing this framework delivers clarity to the methodologies chosen to explore the research aims of this thesis, explicitly stating how and why each method was utilised. This thesis took an exploratory and experimental approach to contributing to the literature on human-machine-interaction. It has been argued that mixed-method approaches are required when researching complex environments (Greene et al., 2001) and sociotechnical environments are inherently complex. Therefore, opting for the use of mixed-methods which required both qualitative and quantitative data was an appropriate approach to take. Multiple methods draw on the strength of both research styles which enable the development of conclusions that can be placed in context of the operational environment. This produces a holistic understanding of the operational use of automated systems. It is argued that triangulation and complementarity that is enabed by using mixed-methods can strengthen real-world research by increasing the validity and depth of research findings. The strength of this thesis lies therefore in the complementary nature of using multiple methods which has increased the depth of understanding gained by the research; further increasing the ecological validity of the work.

NDM research aims to provide practitioners with positive recommendations in order to support their operations and decision making (McAndrew and Gore, 2013). One challenge typical of NDM research and that was encountered by the author of this thesis is gaining the approval for research from several distinct ethical approval boards. It is vital that all psychological research adheres to the appropriate ethical organisational bodies that have a stake in the research. This can however cause delays to the collection of data, in this instance a delay of 8 months was endured due to the requirement to gain ethical approval from the Ministry of Defence's ethical committee.

An additional challenge to NDM research is with developing interpersonal relationships with SME who are vital to enabling conclusions drawn from naturalistic research to be grounded ecologically. There must be 'buy in' from these experts to the research programme which is built through the researcher's consideration of the research aims but more importantly the possible benefits that can be provided to practitioners. For example, study 1 in this thesis had a small response from SME, mainly due to a restricted number of experts and a lack of their availability at the time of the research, but also due to the working relationship between the author of this work and the experts being new. However, study 2 had a 90% response rate of SME asked to complete the questionnaire. This is due to a number of factors: i) buy in from the gatekeepers at Dstl due to having developed a working relationship over a number of years and through clear communication of the merits of the research; ii) the research is highly topical to ongoing inhouse work at Dstl; and iii) it was novel for personnel to be explicitly asked their views on this topic and therefore the research provided them with an outlet to their opinions and views that was not previously available to them. It transpired that the majority of personnel had not been consulted on their views and opinions of automated systems and as they utilise a range of automated systems on a daily basis this is an area that they are well placed and willing to comment on. This thesis has focused on the viewpoints of the end user as the majority of literature on human-machine-interaction has focused on features of the system to provide practical recommendations to system developers. A recognised limitation of this thesis is that system designers were not recruited and consulted on their experiences with working alongside practitioners when developing new automated systems. It would be of interest to explore this experience from both sides of the collaborative process that is required to develop systems that are fit for purpose. It will be through understanding both sides of the table that practical recommendations can be developed for both system designers and end users.

The design of ARCS ensured that a realistic scenario was run to explore the uptake of an automated system during a threat detection task. There is often a tendency for research into human-machine-interaction to use short, specific, and overt scenarios. However, how far this replicates the reality of tasks in the real world is not always clear. The scenarios run in ARCS were long (30 minutes each), presented participants with no clear number of potentially hostile tracks and did not dramatically fluctuate perceived workload. The scenarios were designed to mimic the reality of sitting at an operations desk for long durations without a specific alerted event. Therefore, replicating some of the boredom and monotony of the task that operators face in reality. Although this design increased the ecological validity of the task it did have negative implications on the number of participants that were recruited due to the length of the experiment; reflected in the ideal sample size not being reached. However, the value of exploring automation uptake in an ecologically valid way arguably outweighs the limitations of a small sample size. The design of the experiment further reduced the negative impact of limited sample numbers through utilising a within-subjects approach. This enabled interesting analysis to be conducted which looked at individual performance across both trials. The amount of data collected by ARCS on each participant also strengthens the conclusions that can be made from the experiment. Appendix Seven presents alternative ways in which data collected by ARCS could be explored by future research.

7.6 Future directions

Greene and Caracelli (1997) argue that multiple methods facilitate exploring complex phenomena by providing contextual understanding. However, this depth often comes hand in hand with identifying the complexity that underlies the situation. Therefore, follow through from research may be required in order to ensure that the interpretations of findings can be understood by the intended audience, for instance policy makers (Greene et al., 2001). The remainder of this chapter will present ideas for future directions of research to build from the findings of this thesis.

The main finding from this thesis is the importance of developing and maintaining effective communication and collaboration between serving personnel, researchers and system designers. Therefore, it is recommended that future research continues to explore ways in which to develop and support these inter-disciplinary links, for example the Athena Project. At the time of writing, it is unclear how effective the Athena Project has been for the US Navy in facilitating communication and knowledge exchange between academia, industry and the military. Future research will be required to explore the effectiveness of programmes like the Athena Project and, if found to be effective, similar concepts should look to be developed and utilised worldwide.

The importance of task expertise was also highlighted. For example, identifying the capabilities of personnel to manage the extra workload that implementation of new automated systems can produce whilst still completing their day-to-day tasks. As systems become increasingly technical the nature of task expertise is changing. Personnel are required to develop expertise on the systems they interact with as well as the expertise to manage the task manually should the system fail. It has been evidenced in this thesis that personnel are not commonly consulted in the development of new systems which have a direct impact on their working day tasks. Although not explicitly explored within this thesis the findings of Chapter 4 raise important questions with regards to training. From the authors experience current training procedures vary in terms of what command level, job role and operational systems are being trained. It would therefore be worthwhile to explore the effectiveness of each training path to explore how robust the current training courses are. Effective training is vital to developing adaptive expertise (Brown, 2007; Hutton et al., 2017; Militello and Klein, 2013). Adaptability is a key tenant of expertise and good decision making, it is also necessary for the viability of a sociotechnical system (Naikar, 2018).

Mentioned in the introduction to this thesis was the potential impact decision inertia may have upon automation misuse or disuse. Although this topic was beyond the scope of this thesis, the impact of decision inertia is an interesting area for future research to explore. Within military contexts and as stated by one SME that took part in this research "any decision is better than no decision". However, research has shown that individuals have a tendency to avoid decisions when: faced with non-time bounded situations, working alongside multiple teams, and when dealing with uncertainty in high-accountability environments (Alison et al., 2015; 2010). As automated decision support systems are being introduced into such environments it would be prudent to explore how the application (or disuse) of these systems mitigates against decision inertia or increases its occurrence. Individuals have a tendency to associate less wrong doing with inaction as opposed to action (Eyre et al., 2008a) which can manifest as remaining in the information search phase in order to avoid feelings of regret. Therefore, one possible motivation behind utilising an automated system could be in order to avoid making the difficult decisions. However, the moral and ethical consideration of operating with automation may mitigate this interaction as legally the human operator is accountable should the system make an error. Consequently, does the use of automated systems reduce the occurrence of decision inertia through increasing the accountability felt by personnel when working alongside an automated system as opposed to another human. This is an interesting area of research that would benefit from further exploration.

7.7 Final Conclusions

The aim of this thesis was to explore the uptake of automated systems in the maritime air defence environment. It has contributed to the literature on human-machine-interaction by adding to the limited body of research that has explored automation uptake in operational settings. It achieved this by taking a mixed-methods approach to combine expert knowledge with more traditional experimental measures. Two qualitative questionnaires were developed and distributed to RN SME. With the insight gained from these studies a simulated microworld was developed and tested utilising a student sample. Critically this thesis has identified that new systems are often introduced not fit for purpose due to a disconnect between end user and system designer. This has negative implications on personnel workload, capability development and financial pressures on the organisation. One way in which to address this is by the development and increase in use of immersive simulation environments within research and during training. Therefore, this thesis also provides researchers with a novel microworld - ARCS with which research into the uptake of automation in the maritime environment can be explored in a holistic and ecologically valid way. Additionally, the

ARCS experiment highlighted that task performance is not necessarily always improved by the uptake of automated decision support systems. This finding links back to the responses from personnel, highlighting the crucial importance of supporting collaborative links between system designers and end users. Therefore, this thesis recommends that focus on how to develop and maintain effective links across military organisations, academia and industry is required to ensure that newly designed systems are fit for purpose and do provide the required support to personnel to increase capabilities.

References

Adams, B. and Webb, R. (2002). Trust in small military teams. In *7th International Command and Control Conference*, pages 1–20.

Alison, L. and Crego, J. (2008). *Policing critical incidents*. Willan, Cullumpton, UK.

Alison, L., Eyre, M., and Humann, M. (2010). Losing sight of the golden mean: accountogenic decisions in UK policing. In Mosier, K. and Fischer, U., editors, *Informed by Knowledge: Expert performance in complex situations.*, pages 274 – 294, New York. Psychology Press.

Alison, L., Power, N., van den Heuvel, C., Humann, M., Palasinksi, M., and Crego, J. (2015). Decision inertia: Deciding between least worst outcomes in emergency responses to disasters. *Journal of Occupational and Organizational Psychology*, 88(2):295–321. DOI: 10.1111/joop.12108.

Alison, L., Power, N., van den Heuvel, C., and Waring, S. (2014). A Taxonomy of Endogenous and Exogenous Uncertainty in High-Risk , High-Impact Contexts. *Journal of Applied Psychology*, 100(4):1309.

Alison, L., van den Heuvel, C., Waring, S., Long, A., O'Hara, T., and Crego, J. (2013). Immersive Simulated Learning Environments for Researching Critical Incidents: A knowledge synthesis of the literature and experiences of studying high-risk strategic decision making. *Journal of Cognitive Engineering and Decision Making*, 7(3):255–272. `https://doi.org/10.1177/1555343412468113`.

Anderson, C. J. (2003). The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological bulletin*, 129(1):139–167. DOI: 10.1037/0033-2909.129.1.139.

Aronson, J. (1995). A pragmatic view of thematic analysis. *The Qualitative Report*, 2(1):3. DOI= 10.4135/9781446214565.n17.

Ashleigh, M. J. and Nandhakumar, J. (2007). Trust and technologies: Implications for organizational work practices. *Decision Support Systems*, 43(2):607–617. DOI: 10.1016/j.dss.2005.05.018.

Balfe, N., Sharples, S., and Wilson, J. R. (2015). Impact of automation: Measurement of performance, workload and behaviour in a complex control environment. *Applied ergonomics*, 47:52–64. DOI: 10.1016/j.apergo.2014.08.002.

Bandara, W. (2006). Using NVivo as a research management tool: A case narrative. In *Quality and Impact of Qaulitative Research: Proceedings of the 3rd International Conference on QaulIT*, pages 6–19.

Bandura, A. and Wood, R. (1989). Effect of Perceived Controllability and Performance Standards on Self-Regulation of Complex Decision Making. *Journal of Personality and Social Psychology*, 56(5):805–814. DOI: 10.1037/0022-3514.56.5.805.

Bard, J. F. (1981). An engagement effectiveness model for surface ships. In *Proceedings of the 13th conference on Winter simulation-Volume 1*, pages 83–88. IEEE Press.

Bari, A., Kellermann, T. S., and Studer, B. (2016). Impulsiveness and inhibitory mechanisms. In R, A. J. and J, C., editors, *Neuroimaging Personality, Social Cognition, and Character*, pages 113–136. Elsevier.

Baron, J. (1993). Heuristics and biases in equity judgments: A utilitarian approach. In Mellers, B. A. and Baron, J., editors, *Psychological perspectives on justice: Theory and applications*, pages 109–137. Cambridge University Press, New York.

Barratt, E. (1985). Impulsiveness subtraits: arousal and informaton processing. In Spence, J. and Izard, C., editors, *Motivation, emotion and personality*, pages 137–146. Elsevier Science Publishers, North Holland.

Barrick, M. R. and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x.

Bazeley, P. and Jackson, K. (2013). *Qualitative data analysis with NVivo*. SAGE Publications Ltd, Thousand Oaks, California.

Beck, H. P., McKinney, J. B., Dzindolet, M. T., and Pierce, L. G. (2009). Effects of humanmachine competition on intent errors in a target detection task. *Human factors*, 51(4):477–486. https://doi.org/10.1177/0018720809341746.

Beck, H. R., Dzindolet, M. T., and Pierce, L. G. (2007). Automation usage decisions: controlling intent and appraisal errors in a target detection task. *Human factors*, 49(3):429–437. DOI: 10.1518/001872007X200076.

Bekier, M., Molesworth, B. R., and Williamson, A. (2012). Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety Science*, 50(2):259–265. DOI: 10.1016/j.ssci.2011.08.059.

Benaskeur, A. R., Irandoust, H., Kabanza, F., and Beaudry, E. (2011). *Decision Support Tool for Anti-Ship Missile Defence Operations*. Technical report, Defence Research and Development Canada, Quebec).

Benaskeur, A. R., Kabanza, F., Beaudry, E., and Beaudoin, M. (2008). A probabilistic planner for the combat power management problem. In *Proceedings of the Eighteenth International Conference on Automated Planning and Scheduling*, pages 12–19.

Bigley, G. A. and Roberts, K. H. (2001). The incident command system: High-reliability organizing for complex and volatile task environments. *Academy of Management Journal*, 44(6):1281–1299. https://doi.org/10.5465/3069401.

Binnie, J. (2017). Saudi-led coalition confirms missile attack on UAE ship. Retrieved from: `http://www.janes.com/article/71481/saudi-led-coalition-confirms-missile-attack-on-uae-ship`.

Boland, R., Maheshwari, A., Te'eni, D., Schwartz, D., and Tenkasi, R. (1992). Sharing perspectives in distributed decision making. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'92)*, number November, pages 306–313. DOI: 0-89791-543-7/92/0010/0306.

Boothby, W. H. (2014). *Conflict law: The influence of new weapons technology, human rights and emerging actors*. TMC Asser Press. DOI: 10.1007/978-94-6265-002-2.

Borst, C., Bijsterbosch, V. A., van Paassen, M. M., and Mulder, M. (2017). Ecological interface design : supporting fault diagnosis of automated advice in a supervisory air traffic control task. *Cognition, Technology & Work*, 9(19):1–16. DOI: 10.1007/s10111-017-0438-y.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101. DOI: 10.1191/1478088706qp063oa.

Brehmer, B. (2005). The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control. In *10th International Command and Control Research and Technology Symposium The Future of C2*, pages 365–368.

Brehmer, B. and 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. https://doi.org/10.1016/0747-5632(93)90005-D.

Brown, L. G. F. J. (2007). Applying NDM to world security needs: Perspectives of a US military practitioner. In Hoffman, R, R., editor, *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, pages 377–384, New York. Lawrence Erlbaum Associates.

Brummer, A. and Hirst, D. (1987). US navy ordered to hit back after Exocet kills 28.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*, 6(1):97–113. DOI: 10.1177/1468794106058877.

Byrne, E. (2018). Knowing behavior helps insure models against breakage: A commentary on Kaber's "Issues in human–automation interaction modeling". *Journal of Cognitive Engineering and Decision Making*, 12:67–69. https://doi.org/10.1177/1555343417725669.

Cannon-Bowers, J. A. and Bell, H. H. (1997). Training decision makers for complex environments: Implications of the naturalistic decision making perspective. In Zsambok, C. E. and Klein, G., editors, *Naturalistic decision making*, pages 99–110. Lawrence Erlbaum Associates, New York.

Cannon-Bowers, J. A. and Salas, E. E. (1998). *Making decisions under stress: Implications for individual and team training.* American psychological association.

Carr, N. (2015). *The glass cage: Where automation is taking us.* The Bodley Head, London.

Casner, S. M., Geven, R. W., Recker, M. P., and Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human factors*, 56(8):1506–1516. https://doi.org/10.1177/0018720814535628.

Chen, A. and Bell, P. A. (2016). Microworlds: using computers to understand choices about the use of natural resources. In Gifford, R., editor, *Research Methods for Experimental Psychology*, pages 185–196. John Wiley & Sons Ltd, West Sussex.

Chen, G., Thomas, B., and Wallace, J. C. (2005). A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *Journal of Applied Psychology*, 90(5):827–841. DOI: 10.1037/0021-9010.90.5.827.

Chipman, S. F., Schraagen, J. M., and Shalin, V. L. (2000). Introduction to cognitive task analysis. In Schraagen, J. M., Chipman, S. F., and Shalin, V. L., editors, *Cognitive task analysis*, pages 3–23. Lawrence Erlbaum Associates, New York.

Cobain, I. (2017). Revealed: catalogue of failings that sank falklands warship HMS Sheffield. Retrieved from: https://www.theguardian.com/uk-news/2017/oct/15/revealed-full-story-behind-sinking-of-falklands-warship-hms-sheffield.

Cohen, M. S., Freeman, J. T., and Thompson, B. B. (1997). Integrated critical thinking training and decision support for tactical anti-air warfare. In *Proceedings of the 1997 Command and Control Research and Technology Symposium*, pages 257–268.

Colebank, J. (2008). *Automated Intelligent Agents: Are they trusted members of military teams?* PhD thesis, Naval Postgraduate School.

Costa Jr, P. T. and McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of personality disorders*, 6(4):343–359. https://doi.org/10.1521/pedi.1992.6.4.343.

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

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., and Hanson, W. E. (2003). Advanced mixed methods research designs. In Tashakkori, A. and Teddlie, C., editors, *Handbook of mixed methods in social and behavioral research*, volume 209, page 240. SAGE Publications, Thousand Oaks, California.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(September):319–340. DOI: 10.2307/249008.

de Boisboissel, G. (2017). Is It Sensible to Grant Autonomous Decision-Making to Military Robots of the Future ? In *2017 International Conference on Military Technologies (ICMT)*, pages 738–742.

De Greef, T. E., Arciszewski, H. F. R., and Neerincx, M. A. (2010). Adaptive Automation Based on an Object-Oriented Task Model: Implementation and Evaluation in a Realistic C2 Environment. *Journal of Cognitive Engineering and Decision Making*, 4(2):152–182. DOI: 10.1518/155534310X522851.

Dennis, J. P. and Vander Wal, J. S. (2010). The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cognitive therapy and research*, 34(3):241–253. https://doi.org/10.1007/s10608-009-9276-4.

Denzin, N. K. and Lincoln, Y. S. (1994). *Handbook of qualitative research.* Sage publications Inc, Thousand Oaks, California.

Department of Defense (1987). Formal investigation into the circumstances surrounding the attack on the USS stark. Retrieved from: `http://www.jag.navy.mil/library/investigations/USS%20STARK%20BASIC.pdf`.

Dörner, D. and Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8:1–11. DOI: 10.3389/fpsyg.2017.01153.

Driskell, J. and Johnston, J. (1998). Stress exposure training. In Cannon-Bowers, J. and Salas, E., editors, *Making decisions under stress: implications for individual and team training*, pages 191–217. American Psychological Association, Washington, DC.

Driskell, J. E. and Salas, E. (1991). Group decision making under stress. *Journal of Applied Psychology*, 76(3):473.

Dzindolet, M., Beck, H., Pierce, L., and Dawe, L. (2001). *A framework of automation use.* Technical report, Army Research Laboratory, ARL-TR-2412. DOI: 10.1037/e447142006-001.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6):697–718. DOI: 10.1016/S1071-5819(03)00038-7.

Dzindolet, M. T., Pierce, L. G., and Beck, H. (2010). *An Examination of the Social, Cognitive and Motivational Factors that Affect Automation Reliance.* Ashgate & Gower, Farnham, Surrey, UK. Doi: 10.1017/CBO9781107415324.004.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., and Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors*, 44(1):79–94. DOI: 10.1518/0018720024494856.

Elgafoss, S., Skramstad, T., and Dalberg, V. (2009). Maritime automation, and emergency handling-software: Design, system integration and quality. In *Electric Ship*

Technologies Symposium, 2009. ESTS 2009. IEEE, pages 86–92. IEEE.

Endsley, M. R. (1987). The application of human factors to the development of expert systems for advanced cockpits. In *Proceedings of the Human Factors Society Annual Meeting*, volume 31, pages 1388–1392. SAGE Publications Sage CA: Los Angeles, CA.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting*, volume 32, pages 97–101. SAGE Publications Sage CA: Los Angeles, CA.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1):32–64. DOI:10.1518/001872095779049543.

Endsley, M. R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review Process. In Endsley, M. R. and Garland, D. J., editors, *Situation Awareness Analysis and measurement.*, pages 1–24. Lawrence Erlbaum, Mahwah, NJ.

Endsley, M. R. (2015a). Final Reflections: Situation Awareness Models and Measures. *Journal of Cognitive Engineering and Decision Making*, 9(1):101–111. DOI: 10.1177/1555343415573911.

Endsley, M. R. (2015b). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1):4–32. DOI: 10.1177/1555343415572631.

Endsley, M. R. (2015c). Situation awareness: operationally necessary and scientifically grounded. *Cognition, Technology and Work*, 17(2):163–167. DOI: 10.1007/s10111-015-0323-5.

Endsley, M. R. and Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3):462–492. DOI: 10.1080/001401399185595.

Endsley, M. R. and Kiris, E. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, 37(2):381–394. DOI: 10.1518/001872095779064555.

Entin, E. E. and Serfaty, D. (1999). Adaptive team coordination. *Human factors*, 41(2):312–325. https://doi.org/10.1518/001872099779591196.

Epley, N. and Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science*, 17(4):311–318. `https://doi.org/10.1111/j.1467-9280.2006.01704.x`.

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., Stough, C. O., Canter, K. S., Robles, R., and Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology*, 15(2):160–170. DOI: 10.1016/j.ijchp.2014.12.001.

Eyre, M., Alison, L., Crego, J., and McLean, C. (2008a). Decision inertia: the impact of organisations on critical incident decision-making. In Alison, L. and Crego, J., editors, *Policing Critical Incidents: Leadership and critical incident management*, pages 201–239. Willan, Cullumpton, UK.

Eyre, M., Crego, J., and Alison, L. (2008b). Electronic debriefs and simulations as descriptive methods for defining the critical incident landscape. In Alison, L. and Crego, J., editors, *Policing critical incidents*, pages 24–53. Willan, Cullumpton, UK.

Ezard, J. (2000). Officers dismissd radar warning of exocet attack on HMS Sheffield. Retrieved from: `https://www.theguardian.com/uk/2000/sep/26/falklands.world`.

Faul, F., Erdfelder, E., Lang, A.-G., and 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. https://doi.org/10.3758/BF03193146.

Finch, J. (1987). Research Note The Vignette Technique In Survey Research. *Sociology*, 21(1):105–114. DOI: 0803973233.

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

Foard, T. R. (2000). Theater Air Defense cornerstones. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 21(3):387–392. Retrieved from: http://techdigest.jhuapl.edu/TD/td2103/foard.pdf.

Forcast International (2017). There is something wrong with our ships today - and there is something wrong with our system. `http://www.defense-aerospace.com/articles-view/feature/5/186143/collisions-indicate-systemic-errors-in-us-navigation-procedures.html`.

Fraher, A. (2017). US navy collisions point to the risks of automation on sea, air and land. Retrieved from: http://theconversation.com/us-navy-collisions-point-to-the-risks-of-automation-on-sea-air-and-land-83019.

Frink, D. and Klimosky, R. (1998). Toward a theory of accountability in organizations and human resources management. In Ferris, G., editor, *Research in personnel and human resource management*, pages 1–50. JAI Press, Greenwich, CT.

Gallagher, S. (2002). Experimenting with introspection. *Trends in Cognitive Sciences*, 6(9):374–375.

Ghazizadeh, M., Lee, J. D., and Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*, 14:39–49. DOI: 10.1007/s10111-011-0194-3.

Gonzalez, C., Vanyukov, P., and Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, 21(2):273–286. DOI: 10.1016/j.chb.2004.02.014.

Gore, J., Banks, A., Millward, L., and Kyriakidou, O. (2006). Naturalistic decision making and organizations: Reviewing pragmatic science. *Organization Studies*, 27(7):925–942. `https://doi.org/10.1177/0170840606065701`.

Gore, J. and Conway, G. E. (2016). Modeling and aiding intuition in organizational decision making: A call for bridging academia and practice. *Journal of Applied Research in Memory and Cognition*, 5(3):331–334. `https://doi.org/10.1016/j.jarmac.2016.08.002`.

Greene, J., Benjamin, L., and L, G. (2001). The merits of mixing methods in evaluation. *Evaluation*, 7(1):25–44.

Greene, J. C. and Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New directions for evaluation*, (74):5–17. https://doi.org/10.1002/ev.1068.

Greene, J. C., Caracelli, V. J., and Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis*, 11(3):255–274. `https://doi.org/10.3102/01623737011003255`.

Guest, G., MacQueen, K. M., and Namey, E. E. (2012). *Applied thematic analysis.* SAGE Publications, London.

Hall, A. T., Frink, D. D., and Buckley, M. R. (2017). An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior*, 38(2):204–224. https://doi.org/10.1002/job.2052.

Hammond, J. R. (2017). The forgotten US shootdown of Iranian Airliner Flight 655. Retrieved from: https://www.foreignpolicyjournal.com/2017/07/03/the-forgotten-us-shootdown-of-iranian-airliner-flight-655

Hancock, P., Billings, D., and Oleson, K. (2011). *A Meta-Analysis of Factors Influencing the Development of Human-Robot Trust.* Technical Report December, Army Research Laboratory. Retrieved from: `http://oai.dtic.mil/oai/oai?verb=getRecord{\&}metadataPrefix=html{\&}identifier=ADA556734`.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th annual meeting*, pages 904–908.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52:139–183. DOI: 10.1016/S0166-4115(08)62386-9.

Hawley, J. K., Mares, A. L., and Giammanco, C. A. (2005). *The human side of automation: Lessons for air defense command and control*. Technical Report ARL-TR-368. Technical Report March.

Helldin, T. (2014). *Transparency for Future Semi-Automated Systems: Effects of transparency on operator performance, workload and trust.* PhD thesis. Retrieved from: `http://www.diva-portal.org/smash/record.jsf?pid=diva2:760460`.

Hesse-Biber, S. N. (2010). *Mixed methods research: Merging theory with practice.* Guilford Press.

Hinson, J. M., Jameson, T. L., and Whitney, P. (2002). Somatic markers, working memory, and decision making. *Cognitive, Affective, & Behavioral Neuroscience*, 2(4):341–353. DOI: 10.3758/CABN.2.4.341.

Hoc, J.-M. (2000). From human–machine interaction to human–machine cooperation. *Ergonomics*, 43(7):833–843. `https://doi.org/10.1080/001401300409044`.

Hoff, K. A. and Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3):407–434. DOI: 10.1177/0018720814547570.

Hoffman, R. R. and Militello, L. G. (2012). *Perspectives on cognitive task analysis: Historical origins and modern communities of practice.* Psychology Press, New York.

Hoffman, R. R., Ward, P., Feltovich, P. J., Dibello, L., Fiore, S. M., and Andrews, D. H. (2013). *Accelerated Expertise.* Taylor & Francis, New York.

Holt, J. (1988). Assessing the need for decision support systems. *European Journal of Operational Research*, 37(1):73–82. DOI: 10.1016/0377-2217(88)90281-0.

Hoover, R. S. and Koerber, A. L. (2011). Using NVivo to answer the challenges of qualitative research in professional communication: Benefits and best practices: Tutorial. *IEEE Transactions on Professional Communication*, 54(1):68–82. DOI: 10.1109/TPC.2009.2036896.

Huffington Post (2016). USS mason draws missile fire again off yemen. Retrieved from: `http://www.huffingtonpost.com/entry/uss-mason-missile-yemen_us_5802c72ee4b0162c043c732e`.

Hutchins, S. (1997). Decision-making errors demonstrated by experienced naval officers in a littoral environment. In Zsambok, C. and Klein, G., editors, *Naturalistic Decision Making*, pages 207–216. Lawrence Erlbaum Associates, New Jersey.

Hutton, R., Ward, P., Gore, J., Turner, P., Hoffman, R., Leggatt, A., and Conway, G. (2017). Developing Adaptive Expertise: A synthesis of literature and implications for training. In Gore, J. and Ward, P., editors, *Proceedings of the 13th* *biannual International Conference on Naturalistic Decision Making*, pages 81–86, University of Bath, Bath.

Hutton, R. J., Miller, T. E., and Thordsen, M. L. (2003). Decision-centered design: Leveraging cognitive task analysis in design. In Hollnagel, E., editor, *Handbook of cognitive task design*, pages 383–416. Lawrence Erlbaum Publishers, Mahwah, New Jersey.

Inskip, I. (2002). *Ordeal by Exocet: HMS Glamorgan and the Falklands War, 1982*. Greenhill Books/Lionel Leventhal.

Jamieson, G. A. and Skraaning, G. J. (2017). Levels of automation in human factors models for automation design: Why we might consider throwing the baby out with the bathwater. *Journal of Cognitive Engineering and Decision Making*, 12:42–49. https://doi.org/10.1177/1555343417732856.

Jean, G. V. (2008). Success of simulation-based training is tough to measure. *National Defense*, 93(661):46.

Jenvald, J. and Morin, M. (2004). Simulation-supported live training for emergency response in hazardous environments. *Simulation & Gaming*, 35(3):363–377. https://doi.org/10.1177/1046878104266223.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 24(4):602–611. DOI: 10.2307/2392366.

Johansson, F. (2010). *Evaluating the performance of TEWA systems*. PhD thesis, Örebro universitet.

John, O. P. and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A. and P, J. O., editors, *Handbook of personality: Theory and research (2nd Edition)*, number 1999, pages 102–138. The Guilford Press, New York.

Johnston, J., Cannon-Bowers, J., and Salas, E. (1998). Tactical Decision Making Under Stress (TADMUS): Mapping a program of research to a real world incidentthe USS Vincennes. In *RTO HFM Symposium on Collaborative Crew Performance in Complex Operational Systems*, pages 20–22, Edinburgh, Scotland.

Johnston, J. H., Driskell, J. E., and Salas, E. (1997). Vigilant and hypervigilant decision making. *The Journal of applied psychology*, 82(4):614–622. DOI: 10.1037/0021-9010.82.4.614.

Johnston, N. and McDonald, N. (2017). *Aviation psychology in practice*. Routledge.

Jost, J. T., Glaser, J., Kruglanski, A. W., and Sulloway, F. J. (2003). Political Conservatism as Motivated Social Cognition. *Psychological Bulletin*, 129(3):339–375. DOI: 10.1037/0033-2909.129.3.339.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., and Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, 36(5):447–462. DOI: 10.1016/j.ergon.2006.01.008.

Kaber, D. B., Wright, M. C., Prinzel, L. J., and Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human factors*, 47(4):730–741. DOI: 10.1518/001872005775570989.

Kaempf, G., Klein, G., Thordsen, M., and Wolf, S. (1996). Decision Making in Complex Naval Command and Control Environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(2):220–231. https://doi.org/10.1177/001872089606380204.

Kaempf, G. I., Wolf, S., and Miller, T. E. (1993). Decision making in the AEGIS combat information center. In *Proceedings of the Human Factors and Ergonomics Society*, volume 2, pages 1107–1111. DOI: 10.1177/154193129303701615.

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin Books Ltd., London, England.

Kahneman, D. and Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American psychologist*, 64(6):515. `http://dx.doi.org/10.1037/a0016755`.

Kaplan, S. E., Reneau, J. H., and Whitecotton, S. (2001). The Effects of Predictive Ability Information, Locus of Control, and Decision Maker Involvement on Decision Aid Reliance. *Journal of Behavioral Decision Making*, 14(1):35–50.

Kim, S. S. and Malhotra, N. K. (2005). A longitudinal model of continued IS use: An integrative view of four mechanisms underlying postadoption phenomena. *Management science*, 51(5):741–755.

King's College London (2015). The nature, scale and beneficiaries of research impact: An initial analysis of research excellence framework (ref) 2014 impact case studies.

Kirkman, M. A., Ahmed, M., Albert, A. F., Wilson, M. H., Nandi, D., and Sevdalis, N. (2014). The use of simulation in neurosurgical education and training: a systematic review. *Journal of neurosurgery*, 121(2):228–246.

Klein, G. (2015). Reflections on Applications of naturalistic decision-making. *Journal of Occupational and Organizational Psychology*, 88(2):382–386. DOI: 10.1111/joop.12121.

Klein, G. and Baxter, H. (2006). Cognitive transformation theory: contrasting cognitive and behavioural learning. In *Interservice/Industry Training Systems and Education Conference*, Orlando, Florida.

Klein, G., Calderwood, R., and Clinton-Cirocco, A. (2010). Rapid Decision Making on the Fire Ground: The Original Study Plus a Postscript. *Journal of Cognitive Engineering and Decision Making*, 4(3):186–209. DOI: 10.1518/155534310X12844000801203.

Klein, G., Calderwood, R., and MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, 19(3):462–472. DOI: 10.1109/21.31053.

Klein, G. and Militello, L. (2005). The knowledge audit as a method for cognitive task analysis. In Montgomery, H., Lipshitz, R., and Brehmer, B., editors, *How professionals make decisions*, pages 335–342. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Klein, G., Pliske, R., Crandall, B., and Woods, D. D. (2005). Problem detection. *Cognition, Technology and Work*, 7(1):14–28. DOI: 10.1007/s10111-004-0166-y.

Klein, G., Ross, K. G., Moon, B. M., Klein, Devorah, E., Hoffman, R. R., and Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems*, 18(3):81–85. DOI: 10.1109/MIS.2003.1200735.

Klein, G., Snowden, D., and Pin, C. L. (2011). Anticipatory thinking. In Mosier, K. L. and Fischer, U. M., editors, *Informed by knowledge: Expert performance in complex situations*, pages 235–246. Psychology Press, New York.

Klein, G. A. (1998). *Sources of power: How people make decisions*. MIT press.

Klein, H. (2004). The cultural lens model. In Kaplan, M., editor, *Cultural ergonomics: advances in human performance and cognitive engineering*, pages 249–280. Elsevier, Oxford, UK.

Kobus, D. a., Proctor, S., and Holste, S. (2001). Effects of Experience and Uncertainty during Dynamic Decision Making. *International Journal of Industrial Ergonomics*, 28(5):275–290. DOI: 10.1177/154193120004400147.

Koltai, K., Ho, N., Masequesmay, G., Niedober, D., Skoog, M., Cacanindin, A., Johnson, W., and Lyons, J. (2014). Influence of cultural, organizational, and automation capability on human automation trust: A case study of auto-gcas experimental test pilots. In *HCI-Aero 2014 International Conference on Human-Computer Interaction in Aerospace*. Santa Clara, United States.

Konnikova, M. (2014). The hazards of going on autopilot. https://www.newyorker.com/science/maria-konnikova/hazards-automation.

Kozlowski, S. and DeShon, R. (2004). A psychological fidelity approach to simulated-based training: theory, research and principles. In Schiflett, S., Elliot, L., and Salas, E., editors, *Scaled worlds: Development, validation and application*, pages 75–99. Ashgate, Aldershot, UK.

Kozlowski, S. W. J. (1998). Training and developing adaptive teams: Theory, principles, and research. In Cannon-Bowers, J. and Salas, E., editors, *Decision making under stress: Implications for individual and team training*, pages 115–153. American Psychological Association, Washington, DC.

Krueger, G. (2012). Military engineering psychology- setting the pace for exceptional performance. In Laurence, J. and Matthews, M., editors, *The Oxford Handbook of Military Psychology*, pages 232–240. Oxford University Press, New York.

Kruglanski, A. (1989). *Lay epistemics and human knowledge: Cognitive and motivational biases*. Plenum Press, New York, NY.

Kruglanski, A. W., Orehek, E., Dechesne, M., and Pierro, A. (2010). Lay epistemic theory: The motivational, cognitive, and social aspects of knowledge formation. *Social and Personality Psychology Compass*, 4(10):939–950. https://doi.org/10.1111/j.1751-9004.2010.00308.x.

Kruglanski, A. W., Pierro, A., Higgins, E. T., and Capozza, D. (2007). "on the move" or "staying put": Locomotion, need for closure, and reactions to organizational change. *Journal of Applied Social Psychology*, 37(6):1305–1340. DOI: 10.1111/j.1559-1816.2007.00214.x.

Kruglanski, A. W., Thompson, E. P., Higgins, E. T., Atash, M. N., Pierro, A., Shah, J. Y., and Spiegel, S. (2000). To "do the right thing" or to "just do it": locomotion and assessment as distinct self-regulatory imperatives. *Journal of personality and social psychology*, 79(5):793–815. DOI: 10.1037/0022-3514.79.5.793.

Kruglanski, A. W., Webster, D. M., and Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of personality and social psychology*, 65(5):861.

LaGrone, S. (2017). The attack on USS stark at 30. Retrieved from: https://news.usni.org/2017/05/17/the-attack-uss-stark-at-30.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174. DOI: 10.2307/2529310.

Larken, J. (2002). Military commanderroyal navy. In Flin, R. and Arbuthnot, K., editors, *Incident Command: Tales from the Hot Seat*, pages 105–137. Ashgate Publishing Limited, UK.

Learmount, D. (2019). Lion air lessons. *Aerospace*, 46(1):18–21.

Lee, J. and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270. https://doi.org/10.1080/00140139208967392.

Lee, J. D. and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human factors*, 46(1):50–80. DOI: 10.1518/hfes.46.1.50.30392.

Levine, L. E., Waite, B. M., and Bowman, L. L. (2007). Electronic media use, reading, and academic distractibility in college youth. *CyberPsychology & Behavior*, 10(4):560–566. https://doi.org/10.1089/cpb.2007.9990.

Levs, J. (2012). What caused the cruise ship disaster? `https://edition.cnn.com/2012/01/15/world/europe/italy-cruise-questions/`.

Lewandowsky, S., Mundy, M., and Tan, G. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, 6(2):104.

Li, X., Hess, T. J., and Valacich, J. S. (2008). Why do we trust new technology? a study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, 17(1):39–71. https://doi.org/10.1016/j.jsis.2008.01.001.

Liebhaber, M. and Feher, B. (2002). *Air threat assessment: Research, model, and display guidelines*. Technical report, San Diego. Retrieved from: `http://oai.dtic.mil/oai/oai?verb=getRecord{\&}metadataPrefix=html{\&}identifier=ADA458047`.

Limerick, H., Coyle, D., and Moore, J. W. (2014). The experience of agency in human-computer interactions: a review. *Frontiers in Human Neuroscience*, 8(643):1–10. DOI: 10.3389/fnhum.2014.00643.

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

Lipshitz, R. and Strauss, O. (1997). Coping with Uncertainty: A Naturalistic Decision-Making Analysis. *Organizational Behavior and Human Decision Processes*, 69(2):149–163. `https://doi.org/10.1006/obhd.1997.2679`.

Litt, M. D. (1988). Self-efficacy and perceived control: Cognitive mediators of pain tolerance. *Journal of personality and social psychology*, 54(1):149. http://dx.doi.org/10.1037/0022-3514.54.1.149.

Loomis, J. M., Blascovich, J. J., and Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, 31(4):557–64. DOI: 10.3758/BF03200735.

Lötter, D. and Van Vuuren, J. H. (2014). Implementation challenges associated with a threat evaluation and weapon assignment system. In *Proceedings of the 2014 ORSSA Annual Conference*, pages 27–35.

Machiavelli, N. (2006). *Art of war*. Dover Publications.

Madhavan, P., Wiegmann, D. A., and Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human factors*, 48(2):241–256. https://doi.org/10.1518/001872006777724408.

Martin, M. M. and Anderson, C. M. (1998). The cognitive flexibility scale: Three validity studies. *Communication Reports*, 11(1):1–9. https://doi.org/10.1080/08934219809367680.

McAndrew, C. and 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. DOI: 10.1177/1555343412463922.

McDermott, R. (2017). Russia's new Tsirkon cruise missile. Retrieved from: `http://www.realcleardefense.com/articles/2017/05/04/russias_new_tsirkon_cruise_missile_111307.html`.

McGuirl, J. M. and Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors*, 48(4):656–665. DOI: 10.1518/001872006779166334.

Mendonça, D. (2007). Decision Support for improvisation in response to extreme events: Learning from the response to the 2001 World Trade Centre Attack. *Decision Support Systems*, *3*(43):957–967.

Merritt, S. M. (2011). Affective Processes in Human-Automation Interactions. *Human Factors*, 53(4):356–370. DOI: 10.1177/0018720811411912.

Merritt, S. M., Heimbaugh, H., Lachapell, J., and Lee, D. (2009). I Trust It , but I Don't Know Why : Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors*, 55(3):520–534. DOI: 10.1177/0018720812465081.

Merritt, S. M., Heimbaugh, H., LaChapell, J., and Lee, D. (2013). I trust it, but i dont know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human factors*, 55(3):520–534. DOI: https://doi.org/10.1177/0018720812465081.

Militello, L. and Klein, G. (2013). Decision-centered design. In Lee, J. and Kirlick, A., editors, *The Oxford Handbook of Cognitive Engineering*, pages 261–271. Oxford University Press, Oxford.

Militello, L. G. and Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): a practitioner's. *Ergonomics*, 41(11):1618–1641. https://doi.org/10.1080/001401398186108.

Militello, L. G., Sushereba, C. E., Branlat, M., Bean, R., and Finomore, V. (2015). Designing for military pararescue: Naturalistic decision-making perspective, methods, and frameworks. *Journal of Occupational and Organizational Psychology*, 88(2):251–272. DOI: 10.1111/joop.12114.

Miller, C. A. (2018). The risks of discretization: What is lost in (even good) Levels-of-Automation schemes. *Journal of Cognitive Engineering and Decision Making*, pages 74–76. https://doi.org/10.1177/1555343417726254.

Ministry of Defence (2010). Understanding. Technical Report Joint Doctrine Publication 04, Shrivenham, Wiltshire.

Ministry of Defence (2011). *British Maritime Doctrine.* Technical Report Joint Doctrine Publication 0-10, Shrivenham, Wiltshire.

Ministry of Defence (2014). *UK Defence Doctrine*. Technical Report Joint Doctrine Publication 0-01, (5th Edition), Shrivenham, Wiltshire.

Ministry of Defence (2017). UK Armed Forces Monthly service personnel statistics. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachmentdata/file/659404/20171001SPS.pdf.

Mishler, E. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard educational review*, 60(4):415–443. https://doi.org/10.17763/haer.60.4.n4405243p6635752.

Moray, N., Inagaki, T., and Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of experimental psychology. Applied*, 6(1):44–58. DOI: 10.1037/1076-898X.6.1.44.

Morrison, J. G., Kelly, R. T., Moore, R., and Hutchins, S. G. (1996). *Tactical decision making under stress (TADMUS) decision support system.* Technical report, Naval Postgraduate School.

Morse, J. and Niehaus, L. (2009). *Mixed method design- principles and procedures.* Left Coast Press, Walnut Creek, CA.

Mosier, K. L., Dunbar, M., McDonnell, L., Skitka, L. J., Burdick, M., and Rosenblatt, B. (1998). Automation bias and errors: Are teams better than individuals? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 201–205. SAGE Publications Sage CA: Los Angeles, CA.

Mosier, K. L. and Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In Parasuraman, R. and Mouloua, M., editors, *Automation and human performance: Theory and applications*, page 120. Lawrence Erlbaum Associates.

Moxley, J. H., Anders Ericsson, K., Charness, N., and Krampe, R. T. (2012). The role of intuition and deliberative thinking in experts' superior tactical decision-making. *Cognition*, 124(1):72–78. DOI: 10.1016/j.cognition.2012.03.005.

Muir, B. M. and Moray, N. (1996). Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3):429–460. DOI: 10.1080/00140139608964474.

Murphy, R. and Shields, J. (2012). *Defense science board-task force report–the role of autonomy in DOD systems. Technical Report CD 1172.*

Naikar, N. (2018). Human-Automation Interaction in Self-Organizing Sociotechnical Systems. *Journal of Cognitive Engineering and Decision Making*, 12:62–66. DOI: 10.1177/1555343417731223.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, C. D. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223 – 239. https://doi.org/10.1006/ijhc.1995.1042.

NATO Standarsization Agency (2011). NATO Joint Military Symbology APP-6(C). Retrieved from: `https://www.awl.edu.pl/images/en/APP_6_C.pdf`.

Navarro, J. and Osiurak, F. (2015). When do we use automatic tools rather than doing a task manually? influence of automatic tool speed. *The American journal of psychology*, 128(1):77–88. DOI: 10.5406/amerjpsyc.128.1.0077.

Newson, N. (2016). Future capability of the UK's Armed Forces. Retrieved from: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/LLN-2016-0071.

Nguyen, X. T. (2002). Threat assessment in tactical airborne environments. *Proceedings of the 5th International Conference on Information Fusion, FUSION 2002*, 2:1300–1307.

Onnasch, L. (2015). Crossing the boundaries of automationFunction allocation and reliability. *International Journal of Human-Computer Studies*, 76:12–21. DOI: 10.1016/j.ijhcs.2014.12.004.

Osiurak, F., Wagner, C., Djerbi, S., and Navarro, J. (2013). To do it or to let an automatic tool do it? The priority of control over effort. *Experimental psychology*, 60(6):453–68. DOI: 10.1027/1618-3169/a000219.

Parasuraman, R., de Visser, E., Lin, M.-K., and Greenwood, P. M. (2012). Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *PloS one*, 7(6):e39675. DOI: 10.1371/journal.pone.0039675.

Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. *Human factors*, 52(3):381–410. DOI: 10.1177/0018720810376055.

Parasuraman, R., Molloy, R., and Singh, I. L. (1993). Performance consequences of automation induced "complacency". *The International Journal of Aviation Psychology*, 3(1):1–23. DOI: `10.1207/s15327108ijap0301_1`.

Parasuraman, R. and Riley, V. (1997). Humans and automation : Use , misuse , disuse , abuse. *Human factors*, 39(2):230–253. `https://doi.org/10.1518/001872097778543886`.

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans : a publication of the IEEE Systems, Man, and Cybernetics Society*, 30(3):286–297. DOI: 10.1109/3468.844354.

Parasuraman, R. and Wickens, C. D. (2008). Humans: still vital after all these years of automation. *Human factors*, 50(3):511–520. DOI: 10.1518/001872008X312198.

Pascual, R. and Henderson, S. (1997). Evidence of naturalistic decision making in military command and control. pages 217–226.

Patterson, R., Fournier, L., Pierce, B., Winterbottom, M., and Tripp, L. (2009). Modeling the dynamics of recognition-primed decision making. In *9th International Conference on Naturalistic Decision Making*, London.

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of clinical psychology*, 51(6):768–774.

Pfadenhauer, M. (2009). At Eye-Level: The Expert Interview - a Talk between Expert and Quasi-expert. In Bogner, A., Littig, B., and Menz, W., editors, *Interviewing Experts*, pages 81–97. Palgrave Macmillan.

Powell, J. L. (1991). An attempt at increasing decision rule use in a judgment task. *Organizational Behavior and Human Decision Processes*, 48(1):89–99. https://doi.org/10.1016/0749-5978(91)90007-G.

Power, N. (2017). Extreme teams: Towards a greater understanding of multi-agency teamwork during major emergencies and disasters. *American Psychologist*, 73. http://dx.doi.org/10.1037/amp0000248.

Power, N. S. (2015). *Cognition in Crisis : Decision inertia and failures to take action in multi-agency emergency response command teams*. PhD thesis, University of Liverpool.

Redelmeier, D. A. and Shafir, E. (1995). Medical decision making in situations that offer multiple alternatives. *JAMA*, 273(4):302–305. DOI: 10.1001/jama.1995.03520280048038.

Rice, S. and McCarley, J. S. (2011). Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of Experimental Psychology: Applied*, 17(4):320–331. DOI: 10.1037/a0024243.

Riffenburgh, R. (1991). *Tactical Decision Making Under Stress (TADMUS) Program Study Initial Design*. Technical Report November, Naval Ocean Systems Center, San Diego.

Roberts, A., Stanton, N., and Fay, D. (2015). The command team experimental testbed stage 1: design and build of a submarine command room simulator. *Procedia Manufacturing*, 3:2800–2807. https://doi.org/10.1016/j.promfg.2015.07.745.

Robertson, D. W., Martin, D. K., and Singer, P. A. (2003). Interdisciplinary research: putting the methods under the microscope. *BMC Medical Research Methodology*, 3(1):20. https://doi.org/10.1186/1471-2288-3-20.

Ross, K. G., Klein, G. A., Thunholm, P., Schmitt, J. F., and Baxter, H. C. (2004). The recognition-primed decision model. Technical report, Army combined arms center Fort Leavenworth KS Military Review.

Ross, K. G., Shafer, J. L., and Klein, G. (2006). Professional judgments and "naturalistic decision making. In Ericsson, K. A., Charness, N., Feltovich, P. J., and R, H. R., editors, *The Cambridge handbook of expertise and expert performance*, pages 403–419. Cambridge University Press, New York.

Roth, E. M. and Pritchett, A. R. (2017). Preface to the special issue on advancing models of human–automation interaction. *Journal of Cognitive Engineering and*

Decision Making, 1:3–6. https://doi.org/10.1177/1555343417749192.

Rotliman, A. J. and Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin*, 24(10):1053–1064. https://doi.org/10.1177/01461672982410003.

Röttger, S., Bali, K., and Manzey, D. (2009). Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics*, 52(5):512–523. https://doi.org/10.1080/00140130802379129.

Rouse, W. (1991). *Design for success*. John Wiley, New York, NY.

Rouse, W. B. and Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological bulletin*, 100(3):349.

Roux, J. N. and 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. http://dx.doi.org/10.5784/23-2-54.

Roy, J., Paradis, S., and Allouche, M. (2002). Threat Evaluation for impact assessment in situation analysis systems. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 4729, pages 329–341. DOI: 10.1117/12.477618.

Royal Navy (n.d.). *Future Navy Vision*. Retrieved from= `https://www.royalnavy.mod.uk/About-the-Royal-Navy/~/media/Files/Navy-PDFs/About-the-Royal-Navy/Future%20Navy%20Vision.pdf`.

Royal Navy. (n.d.). *Maritime Warfare School*. `https://www.royalnavy.mod.uk/our-organisation/where-we-are/training-establishments/hms-collingwood/maritime-warfare-school`.

Salas, E. and Cannon-Bowers, J. A. (1997). A framework for developing team performance measures in training. In Brannick, M. T., Salas, E., and W, P. C., editors, *Team performance assessment and measurement: Theory, Methods and Applications*, pages 57–74. Psychology Press, New York.

Sarter, N., Woods, D., and Billings, C. (1997). Automation surprises. In Salvendy, G., editor, *Handbook of Human Factors and Ergonomics*, pages 1926–1943. Wiley, New York.

Schraagen, J. M., Klein, G., and Hoffman, R. R. (2008). *The macrocognition framework of naturalistic decision making*. Ashgate Publishing Limited Aldershot.

Schraagen, J. M. and van de Ven, J. G. M. (2008). Improving Decision Making in Crisis Response Through Critical Thinking Support. *Journal of Cognitive Engineering and Decision Making*, 2(4):311–327. DOI: 10.1518/155534308X377801.

Schwarz, J., Fuchs, S., and Flemisch, F. (2014). Towards a More Holistic View on User State Assessment in Adaptive Human-Computer Interaction. In *2014 IEEE International Conference on System, Man, and Cybernetics*, pages 1228–1234, San Diego.

Schwarz, N. and Vaughn, L. A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. In Gilovich, T., Griffin, D., and Kahneman, D., editors, *Heuristics and biases: The psychology of intuitive judgment*, pages 103–119. Cambridge University Press.

Scott, R. (2017). Attacks reaffirm need for force protection. Retrieved from: `http://www.janes.com/article/67874/attacks-reaffirm-need-for-force-protection-idex17d1`.

Seong, Y. and Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, 38(7-8):608–625. DOI: 10.1016/j.ergon.2008.01.007.

Serfaty, D., Entin, E., and Johnston, J. (1998). Team coordination training. In Cannon-Bowers, J. and Salas, E., editors, *Decision making under stress: Implications for individual and team training*, pages 221–245. American Psychological Association, Washington, DC.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition*, 50(1-3):403–430. https://doi.org/10.1016/0010-0277(94)90038-8.

Shortland, N. (2017). *Conflict: Sacred values, decision inertia and the psychology of choice in military decision-making.* PhD thesis, University of Liverpool.

Shortland, N. and Alison, L. (2015). Least worst decisions 'in extremis' (manuscript in preparation).

Skitka, L., Mosier, K., and Burdick, M. (2000). Accountability and automation bias. *International Journal of Human Computer Studies*, 52(4):701–717. DOI: 10.1006/ijhc.1999.0349.

St. John, M., Smallman, H. S., Manes, D. I., Feher, B. a., and Morrison, J. G. (2005). Heuristic Automation for Decluttering Tactical Displays. *Human Factors*, 47(3):509–525. DOI: 10.1518/001872005774860014.

Stanton, N. A. and Ashleigh, M. (2000). A field study of team working in a new human supervisory control system. *Ergonomics*, 43(8):1190–1209. https://doi.org/10.1080/00140130050084941.

Stanton, N. A. and Roberts, A. P. J. (2017). Examining social , information and task networks in submarine command and control. *IEEE Transactions on Human-Machine Systems*, 48(3):252–265. DOI: 10.1109/THMS.2017.2720659.

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

Stowers, K., Oglesby, J., Sonesh, S., Leyva, K., Iwig, C., and Salas, E. (2017). A framework to guide the assessment of human–machine systems. *Human factors*, 59(2):172–188. https://doi.org/10.1177/0018720817695077.

Strauch, B. (2017). The automation-by-expertise-by-training interaction: Why automation-related accidents continue to occur in sociotechnical systems. *Human factors*, 59(2):204–228. https://doi.org/10.1177/0018720816665459.

Sweller, J., Ayres, P., and Kalyuga, S. (2011). Intrinsic and extraneous cognitive load. In Sweller, J., Ayres, P., and Kalyuga, S., editors, *Cognitive Load Theory*, pages 57–69. Springer, New York.

Szalma, J. L. and Taylor, G. S. (2011). Individual Differences in Response to Automation : The Five Factor Model of Personality. *Journal of Experimental Psychology: Applied*, 17(2):71–96. DOI: 10.1037/a0024170.

Taylor, S. E. (1982). The availability bias in social perception and interaction. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgement under uncertainty: heuristics and biases*, pages 190–200. Cambridge University Press.

"The Athena Project" (2013). The athena project. Retrieved from: `http://www.secnav.navy.mil/innovation/HTML_Pages/2015/05/AthenaProject7.htm`.

Tian, N., Fleurant, A., P.D., W., and S.T., W. (2017). Trends in world military expenditure, 2016. Retrieved from: https://www.sipri.org/sites/default/files/Trends-world-military-expenditure-2016.pdf.

Todd, P. and Benbasat, I. (1994). The Influence of Decision Aids on Choice Strategies - An Experimental-Analysis of the Role of Cognitive Effort. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(4):537–547. DOI: 10.1006/obhd.1994.1074.

Tolcott, M. A. and Holt, V. E. (1987). Impact and potential of decision research on decision aiding. Technical report, American Psychological Association Inc, Washington DC.

Trevino, L. K. (1986). Ethical Decision Making in Organizations : A Person-Situation Interactionist Model. *The Academy of Management Review*, 11(3):601–617. https://doi.org/10.5465/amr.1986.4306235.

Tversky, A. and Khaneman, D. (1973). Availability: a heuristic for judging frequency and probaility. *Cognitive Psychology*, 5(2):207–232. https://doi.org/10.1016/0010-0285(73)90033-9.

van den Heuvel, C., Alison, L., and Crego, J. (2012). How uncertainty and Accountability can derail strategic 'Save life' decisions. *Journal of Behavioral Decision Making*, 25(2):165–187. DOI: 10.1002/bdm.

van Teijlingen, E. R. and Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, (35).

Von Clausewitz, C. (1832). *On war*. Penguin Books, London.

Wang, L., Jamieson, G. a., and Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human factors*, 51(3):281–291. DOI: 10.1177/0018720809338842.

Waring, S., Alison, L., Carter, G., Barrett-Pink, C., Humann, M., Swan, L., and Zilinsky, T. (2018). Information sharing in interteam responses to disaster. *Journal of Occupational and Organizational Psychology*. https://doi.org/10.1111/joop.12217.

Whitney, P., Jameson, T., and Hinson, J. M. (2004). Impulsiveness and executive control of working memory. *Personality and Individual Differences*, 37(2):417–428. https://doi.org/10.1016/j.paid.2003.09.013.

Wickens, C. D. and Hollands, J. G. (2000). Attention, time-sharing, and workload. *Engineering psychology and human performance*, 3:439–479.

Wiener, E. L. and Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10):995–1011. https://doi.org/10.1080/00140138008924809.

Williams, M. B. S. (2010). Heuristics and biases in Military Decision Making. *Military Review*, (October):40–52. Retrieved from: http://www.dtic.mil/dtic/tr/fulltext/u2/a532527.pdf.

Wong, W. (2003). Critical decision method data analysis. In Daiper, D. and Stanton, N. A., editors, *The handbook of task analysis for human-computer interaction*, pages 327–346. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Woods, D. D. (2016). The risks of autonomy: Doyles catch. *Journal of Cognitive Engineering and Decision Making*, 10(2):131–133. https://doi.org/10.1177/1555343416653562.

Yearta, S. K., Maitlis, S., and Briner, R. B. (1995). An exploratory study of goal setting in theory and practice: A motivational technique that works? *Journal of Occupational and Organizational Psychology*, 68(3):237–252. https://doi.org/10.1111/j.2044-8325.1995.tb00584.x.

Zsambok, C. E. and Klein, G. (1997). *Naturalistic decision making*. Lawrence Erlbaum Associates, New York.

Appendices

A Appendix One

Stage 1: Vignette Survey

1.1 Participant information and consent form

What is the research? You have been asked to take part in research that will examine decisions made within a ships operations room. In particular we are looking at understanding situations specific to air defence. All the data gathered from these scenarios and questions will examine your opinions on the decision process that occurs within a ships operations room, from how situational awareness is developed to when defensive actions are taken. Participation is completely anonymous and confidential. Your participation is completely voluntary and you are also free to withdraw from the research at any time should you decide you no longer wish to take part. The results from this questionnaire will be used to develop training aides and software improvements to aid the decision-making process, with the aim to reduce the potential for indecision to occur.

1. I understand that my participation is voluntary and that I am free to withdraw at any time.

2. I understand that none of my personal details will be recorded and that my responses are anonymous.

3. I understand that in order to take part in the study I must be currently (or have been in the past) a member of the Royal Navy and that I have experience of working within a ships operations room.

4. I understand that data supplied by me can be removed at my request at any time after my participation.

5. I understand that all answers to the questionnaire must be at the unclassified level and I must not provide any information that is classified or sensitive.

I understand the above statements and agree to continue with this questionnaire

- ❍ I give consent to take part in this questionnaire (4)

1.2 Stage 2 Vignette Questionnaire

Q1 Date of Birth:

Q2 Gender

- ❍ Male (1)
- ❍ Female (2)

Q3 Rank:

Q4 Time spent serving (years and months):

Q5 Role duties and responsibilities you have conducted (e.g. FC, PWO etc.):

Please read through the scenario below, following will be 4 vignettes, each consisting of several questions. Please write your responses in the boxes provided.

You are in a T45 deployed, as part of a coalition maritime force, in response to a developing regional crisis. There have been no attacks yet on coalition forces however the designated hostile nation maintains aggressive rhetoric and its armed forces are at a high state of readiness. The Political Policy Indicator (PPI) is 'Yankee' - maintain the status quo (significant use of force to support specific objectives may be authorised). The Joint Operating Area (JOA) has been divided into two sectors with maritime Task Groups operating in each sector. Within its Task Group the T45 has been assigned the duty of Sector Air Defence Commander responsible for Battlespace Management within the assigned sector. This includes air identification and maintenance of the Recognised Air Picture, asset management, airspace de-confliction and control of air assets. The T45 is operating in international waters off the coastline of a potentially hostile nation. Hostile nation territorial waters (TTW) and territorial airspace (TTAS) extend out to 12nm from the coastline.

Q1 Situation 1 - A number of active civil airways run either through or adjacent to the assigned sector of responsibility. A daily Air Tasking Order is being issued for friendly military air traffic operating in the JOA. T45 tasked to monitor air activity over the designated hostile nation and in particular any surveillance aircraft operating over the sea.

Q1.1 Explain your thought process and the information used to develop shared situational awareness

Q1.2 Which information is most important to your decision making process?

Q1.3 Please answer the following questions

	List your two key priorities in response to this situation?	Please list the actions you will take to achieve each priority	Please indicate why these actions will achieve each specific priority
	Priorities (1)	What actions taken (1)	Why these actions (1)
Priority (1)			
Priority (2)			

2

Q2 Situation 2- Air contact detected in the vicinity of hostile airbase, transits at 280 knots towards coastline then flies a 30nm barrier at 14,000 feet over the sea but inside TTAS. An airborne maritime search radar is detected on same bearing. Sometime later a new contact detected leaving the hostile coast at 100 knots/ 3,000 feet flies on a direct approach towards the T45.

Q2.1 Explain your thought process and the information used to develop shared situational awareness

Q2.2 Which information is most important to your decision making process?

Q2.3 Please answer the following questions

	List your two key priorities in response to this situation?	Please list the actions you will take to achieve each priority	Please indicate why these actions will achieve each specific priority
	Priorities (1)	What actions taken (1)	Why these actions (1)
Priority (1)			
Priority (2)			

Q3 Situation 3- A hostile MPA has been identified and tracked remaining inside own territorial airspace. Two new air contacts detected flying in formation at 450 knots/ 20,000 feet initially operating 50nm from T45 in territorial airspace but then descend to 500 feet and accelerate towards T45. Air contacts climb to 5000 feet when 30nm from T45 and turn on search radars. Initially no response to warnings but turn away at 15nm and establish radio contact.

Q3.1 Explain your thought process and the information used to develop shared situational awareness

Q3.2 Which information is most important to your decision making process?

Q3.3 Please answer the following questions

	List your two key priorities in response to this situation?	Please list the actions you will take to achieve each priority	Please indicate why these actions will achieve each specific priority
	Priorities (1)	What actions taken (1)	Why these actions (1)
Priority (1)			
Priority (2)			

Q4 Situation 4- Hostile maritime force at sea and closing coalition Task Group, will soon be within range to use their Surface to Surface Guided weapons. Mobile coastal defence missile batteries have been moved to coastal locations putting coalition Task Group inside their engagement envelope. Large scale air activity detected over hostile territory, formations of aircraft marshalling close to the coastline.

Q4.1 Explain your thought process and the information used to develop shared situational awareness?

Q4.2 Which information is most important to your decision making process?

Q4.3 Please answer the following questions

	List your two key priorities in response to this situation?	Please list the actions you will take to achieve each priority	Please indicate why these actions will achieve each specific priority
	Priorities (1)	What actions taken (1)	Why these actions (1)
Priority (1)			
Priority (2)			

4

B Appendix Two

Stage 2 Study Information and Questionnaire

1.1 Participant Information Sheet

Exploring cognitions behind automated use

MoDREC Application No : 785/MoDREC/16

Invitation to Take Part

We would like to invite you to take part in a research study exploring cognitions that may underpin automation use. This research is part a PhD project funded by the Defence Science and Technology Laboratory's (Dstl) National PhD Scheme. 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 us if you would like more information or if there is anything that you do not understand.

What is the Purpose of the Research ?

The aim of this study is to explore the cognitive processes behind automation usage decisions in critical environments for example during the air defence task.

Who is Doing This Research ?

This research is being conducted by Miss Chloe Barrett-Pink as part of her PhD project supervised by Prof. Simon Maskell and Prof. Laurence Alison. The technical partner for this research is Mr Robert Walden. This research is funded by MOD via Dstl's National PhD Scheme.

Why Have I Been Invited to Take Part ?

You have been invited to take part in this research as you have been identified as having the appropriate experience and/or training relating to above water warfare operations. This research aims to help further our understanding of automation usage decisions.

Do I Have to Take Part ?

No, your participation is completely voluntary and you are free to withdraw from the research at any time should you decide you no longer wish to take part without giving a reason. This will not affect your training, or your military career, in any way.

What Will I Be Asked to Do ?

You will be asked to complete a series of questionnaires. Included in this will be several questions that relate to your personal experience of using automated tools/systems during your career. Please make sure any experiences described are generic and avoid any details that would remove the anonymity of your answer.

All results collected will be anonymised and will be used for this study and future research within this project. If you wish to remain updated with the findings of this study and the overall project that your participation will aid, please contact the researcher.

What are the Benefits of Taking Part ?

There are no direct benefits from taking part. However, the results from this study will be used to further our understanding of the influence cognitive factors have upon the decision to use automated decision support systems.

What are the Possible Disadvantages and Risks of Taking Part ?

There are no significant risks or disadvantages anticipated from taking part.

Can I Withdraw from the Research and What Will Happen If I Withdraw ?

Yes, you are free to withdraw from the research at any time should you no longer wish to take part and you do not have to give a reason for doing so. This will not affect your military career, in any way. Any data that has been collected up to the point at which you choose to withdraw will be removed from analysis and destroyed.

Are There Any Expenses and Payments Which I Will Get ?

No

Will My Taking Part or Not Taking Part Affect My Service Career ?

No, this research project will not impact your service career in any way

Whom Do I Contact If I Have Any Questions?

If you would like any more information on the researcher and what the study entails please email Prof. Simon Maskell (smaskell@liverpool.ac.uk) or Chloe Barrett-Pink (c.barrett-pink@liverpool.ac.uk).

Whom Do I Contact If I Have a Complaint?

In accordance with standard practice at the University of Liverpool, if you have any complaints please contact the University Ethics Committee (ethics@liverpool.ac.uk)

What Happens If I Suffer Any Harm ?

In the unlikely event that you suffer any harm as a direct result of taking part in this study, you can apply for compensation under the MoD's 'No-Fault Compensation Scheme.

Will My Records Be Kept Confidential ?

All the data collected will be anonymised and stored in a locked filing cabinet and/or on a password protected storage device during the course of this research. Following completion all consent forms will be forwarded to the MoDREC Secretariant to be stored for 50 years in accordance with extant UK legislation and MoD policy.

Who is Organising and Funding the Research ?

The research is part of a Dstl PhD funded studentship

Who Has Reviewed the Study ?

This study has been reviewed and given favourable opinion by the Ministry of Defence Research Ethics Committee (MoDREC).

This study has also been reviewed and given favourable opinion by the University of Liverpool Ethics Committee.

Further Information and Contact Details

Name : Miss Chloe Barrett-Pink

Address : University of Liverpool, School of Psychological Sciences, Bedford Street South, L69 7ZA
Tel No : +44(0)151 794 3936
E-mail : c.barrett-pink@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.

Thank you for taking the time to read this information sheet. If you would like to take part in this research then please read and sign the attached consent form.

1.2 Participant consent Form

Exploring cognitions behind automation use

MoDREC Reference: 785/MoDREC/16

Thank you for considering taking part in this research. The aim of this research is to explore the potential link between certain cognitive traits and attitudes towards automation use. This survey consists of 8 sections each designed to explore specific cognitions and their association to acceptance and use of automated decision support tools. You will be asked to complete a series of questionnaires. All results collected will be anonymised and will be used for this study and future research within this project. If you wish to remain updated with the findings of this study and the overall project that your participation will aid, please contact the researcher. The survey should take around 20 minutes, all responses are anonymous.

Consent

Please read each statement and check the box if you agree. By checking this box and continuing with this survey you are consenting to take part in this research.

1. The nature, aims and risks of the research have been explained to me. I have read and understood the Information for Participants and understand what is expected of me. All my questions have been answered fully to my satisfaction. ☐

2. I understand that if I decide at any time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately without having to give a reason. I also understand that I may withdraw from it at any time, and that in neither case with this be held against be in subsequent dealings with the Ministry of Defence. ☐

[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).

3. I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998. ☐

4. I agree to volunteer as a participant for the study described in the information sheet and give full consent. ☐

Participant's Statement:

I **…………………………………………………**

agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information for Participants about the project, and understand what the research study involves.

Signed: ________________________________ **Date :** ____________

Investigator's Statement:

I Chloë Barrett-Pink confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the Participant

Signed: Date:

Authorising Signatures

The information supplied above is to the best of my knowledge and belief accurate. I clearly understand my obligations and the rights of research participants, particularly concerning recruitment of participants and obtaining valid consent.

Signature of Chief Investigator

…………………………………… **Date:**

Name and Contact Details of Chief Investigator:
Prof. Simon Maskell
Email: smaskell@liverpool.ac.uk
Telephone: +44(0)151 794 4573

Student Researcher:
Name: Chloë Barrett-Pink
Email: hlcbarre@liverpool.ac.uk

1.3 Stage 2 - Questionnaire

1. Section 1 (of 8)

Please provide the following information about yourself by writing your answer in the space provided.

1. Gender ……………………………
2. What is your age in years? ………………………………..
3. What is your current job title? ……………………………………………………………………………………
4. How long have you held this role? (in years and months)

   ………
5. What previous job roles have you held? (please provide length of time spent at each)

   ………
6. In total, over the course of your career, how long have you spent at sea?

   ………
7. How long has it been since you were last at sea?

   ………

2. Section 2 (of 8)

Please ensure that any descriptions be generic and avoid sensitive details that would remove the anonymity of your answers.

1. In your opinion, do you see automated tools/systems having a role in future naval operations?

2. Where do you see such tools/systems having the most benefit and why?

3. During your time spent at sea or during training, how often did you interact with and utilise automated tools or systems? (i.e. daily, weekly, monthly, once etc)

……..

4. What were the tools you used and how did they aid your operations?

5. Have you ever been consulted in the development of new tools/systems prior to their release into operational use?

6. What are your views on the consultation of current and future personnel during the design and development stages of new automated tools/systems?

3. Section 3 (of 8)

Please use the scale below to indicate the extent to which you agree or disagree with the following statements:

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Neutral*	*Moderately Agree*	*Agree*	*Strongly Agree*
1. I am good at "sizing up" situations	☐	☐	☐	☐	☐	☐	☐
2. I have a hard time making decisions when faced with difficult situations	☐	☐	☐	☐	☐	☐	☐
3. I consider multiple options before making a decision	☐	☐	☐	☐	☐	☐	☐
4. When I encounter difficult situations, I feel like I am losing control	☐	☐	☐	☐	☐	☐	☐
5. I like to look at difficult situations from many different angles	☐	☐	☐	☐	☐	☐	☐
6. I seek additional information not immediately available before attributing causes to behaviour	☐	☐	☐	☐	☐	☐	☐
7. When encountering difficult situations, I become so stressed that I cannot think of a way to resolve the situation	☐	☐	☐	☐	☐	☐	☐
8. I try to think about things from another person's point of view	☐	☐	☐	☐	☐	☐	☐
9. I find it troublesome that there are so many different ways to deal with difficult situations	☐	☐	☐	☐	☐	☐	☐
10. I am good at putting myself in others' shoes	☐	☐	☐	☐	☐	☐	☐
11. When I encounter difficult situations, I just don't know what to do	☐	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Neutral*	*Moderately Agree*	*Agree*	*Strongly Agree*
12. It is important to look at difficult situations from many angles	☐	☐	☐	☐	☐	☐	☐
13. When in difficult situations, I consider multiple options before deciding how to behave	☐	☐	☐	☐	☐	☐	☐
14. I often look at a situation from different view points	☐	☐	☐	☐	☐	☐	☐
15. I am capable of overcoming the difficulties in life that I face	☐	☐	☐	☐	☐	☐	☐
16. I consider all the available facts and information when attributing causes to behaviour	☐	☐	☐	☐	☐	☐	☐
17. I feel I have no power to change things in difficult situations	☐	☐	☐	☐	☐	☐	☐
18. When I encounter difficult situations, I stop and try to think of several ways to resolve it	☐	☐	☐	☐	☐	☐	☐
19. I can think of more than one way to resolve a difficult situation I'm confronted with	☐	☐	☐	☐	☐	☐	☐
20. I consider multiple options before responding to difficult situations	☐	☐	☐	☐	☐	☐	☐

4. Section 4 (of 8)

Read each of the following statements and decide how much you agree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	*Moderately Disagree*	*Slightly Disagree*	*Slightly Agree*	*Moderately Agree*	*Strongly Agree*
1. I think that having clear rules and order at work is essential for success	☐	☐	☐	☐	☐	☐
2. Even after I've made up my mind about something, I am always eager to consider a different option	☐	☐	☐	☐	☐	☐
3. I don't like situations that are uncertain	☐	☐	☐	☐	☐	☐
4. I dislike questions which could be answered in many different ways	☐	☐	☐	☐	☐	☐
5. I like to have friends who are unpredictable	☐	☐	☐	☐	☐	☐
6. I find that a well-ordered life with regular hours suits my temperament	☐	☐	☐	☐	☐	☐
7. I enjoy the uncertainty of going into a new situation without knowing what might happen	☐	☐	☐	☐	☐	☐
8. When dining out, I like to go to places where I have been before so that I know what to expect	☐	☐	☐	☐	☐	☐
9. I feel uncomfortable when I don't understand the reason why an event occurred in my life	☐	☐	☐	☐	☐	☐
10. I feel irritated when one person disagrees with what everyone else in a group believes	☐	☐	☐	☐	☐	☐
11. I hate to change plans at the last minute	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Moderately Disagree*	*Slightly Disagree*	*Slightly Agree*	*Moderately Agree*	*Strongly Agree*
12. I would describe myself as indecisive	☐	☐	☐	☐	☐	☐
13. When I go shopping, I have difficulty deciding exactly what it is I want	☐	☐	☐	☐	☐	☐
14. When faced with a problem I usually see the one best solution very quickly	☐	☐	☐	☐	☐	☐
15. When I am confused about an important issue, I feel very upset	☐	☐	☐	☐	☐	☐
16. I tend to put off making important decisions until the last possible moment	☐	☐	☐	☐	☐	☐
17. I usually make important decisions quickly and confidently	☐	☐	☐	☐	☐	☐
18. I have never been late for an appointment or work	☐	☐	☐	☐	☐	☐
19. I think it is fun to change my plans at the last moment	☐	☐	☐	☐	☐	☐
20. My personal space is usually messy and disorganised	☐	☐	☐	☐	☐	☐
21. In most social conflicts, I can easily see which side is right and which is wrong	☐	☐	☐	☐	☐	☐
22. I have never known someone I did not like	☐	☐	☐	☐	☐	☐
23. I tend to struggle with most decisions	☐	☐	☐	☐	☐	☐
24. I believe orderliness and organisation are among the most important characteristics of a good student	☐	☐	☐	☐	☐	☐
25. When considering most conflict situations, I can usually see how both sides could be right	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Moderately Disagree*	*Slightly Disagree*	*Slightly Agree*	*Moderately Agree*	*Strongly Agree*
26. I don't like to be with people who are capable of unexpected actions	☐	☐	☐	☐	☐	☐
27. I prefer to socialise with familiar friends because I know what to expect from them	☐	☐	☐	☐	☐	☐
28. I think that I would learn best in a class that lacks clearly stated objectives and requirement	☐	☐	☐	☐	☐	☐
29. When thinking about a problem, I consider as many different opinions on the issue as possible	☐	☐	☐	☐	☐	☐
30. I don't like to go into a situation without knowing what I can expect from it	☐	☐	☐	☐	☐	☐
31. I like to know what people are thinking all the time	☐	☐	☐	☐	☐	☐
32. I dislike it when a person's statement could mean many different things	☐	☐	☐	☐	☐	☐
33. It's annoying to listen to someone who cannot seem to make up his or her mind	☐	☐	☐	☐	☐	☐
34. I find that establishing a consistent routine enables me to enjoy life more	☐	☐	☐	☐	☐	☐
35. I enjoy having a clear and structured mode of life	☐	☐	☐	☐	☐	☐
36. I prefer interacting with people whose opinions are very different from my own	☐	☐	☐	☐	☐	☐
37. I like to have a plan for everything and a place for everything	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Moderately Disagree*	*Slightly Disagree*	*Slightly Agree*	*Moderately Agree*	*Strongly Agree*
38. I feel uncomfortable when someone's meaning or intention is unclear to me	☐	☐	☐	☐	☐	☐
39. I believe that one should never engage in leisure activities	☐	☐	☐	☐	☐	☐
40. When trying to solve a problem I often see so many possible options that it's confusing	☐	☐	☐	☐	☐	☐
41. I always see many possible solutions to problems I face	☐	☐	☐	☐	☐	☐
42. I'd rather know bad news than stay in a state of uncertainty	☐	☐	☐	☐	☐	☐
43. I feel that there is no such thing as an honest mistake	☐	☐	☐	☐	☐	☐
44. I do not usually consult many different options before forming my own view	☐	☐	☐	☐	☐	☐
45. I dislike unpredictable situations	☐	☐	☐	☐	☐	☐
46. I have never hurt another person's feelings	☐	☐	☐	☐	☐	☐
47. I dislike the routine aspects of my work	☐	☐	☐	☐	☐	☐

5. Section 5.1 (of 8)

Read each of the following statements and decide how much you agree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Moderately Agree*	*Agree*	*Strongly Agree*
6. I don't mind doing things even if they involve extra effort	☐	☐	☐	☐	☐	☐
7. I am a "workaholic"	☐	☐	☐	☐	☐	☐
8. I feel excited just before I am about to reach a goal	☐	☐	☐	☐	☐	☐
9. I enjoy actively doing things, more than just watching and observing	☐	☐	☐	☐	☐	☐
10. I am a "doer"	☐	☐	☐	☐	☐	☐
11. When I finish one project, I often wait awhile before getting started on a new one	☐	☐	☐	☐	☐	☐
12. When I decide to do something, I can't wait to get started	☐	☐	☐	☐	☐	☐
13. By the time I accomplish a task, I already have the next one in mind	☐	☐	☐	☐	☐	☐
14. I am a "low energy" person	☐	☐	☐	☐	☐	☐
15. Most of the time my thoughts are occupied with the task I wish to accomplish	☐	☐	☐	☐	☐	☐
16. When I get started on something, I usually persevere until I finish it	☐	☐	☐	☐	☐	☐
17. I am a "go-getter"	☐	☐	☐	☐	☐	☐

Section 5.2 (of 8)

Read each of the following statements and decide how much you agree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Moderately Agree*	*Agree*	*Strongly Agree*
1. I never evaluate my social interactions with others after they occur	☐	☐	☐	☐	☐	☐
2. I spend a great deal of time taking inventory of my positive and negative characteristics	☐	☐	☐	☐	☐	☐
3. I like evaluating other people's plans	☐	☐	☐	☐	☐	☐
4. I often compare myself with other people	☐	☐	☐	☐	☐	☐
5. I don't spend much time thinking about ways others could improve themselves	☐	☐	☐	☐	☐	☐
6. I often critique work done by myself or others	☐	☐	☐	☐	☐	☐
7. I often feel that I am being evaluated by others	☐	☐	☐	☐	☐	☐
8. I am a critical person	☐	☐	☐	☐	☐	☐
9. I am very self-critical and self-conscious about what I am saying	☐	☐	☐	☐	☐	☐
10. I often think that other people's choices and decisions are wrong	☐	☐	☐	☐	☐	☐
11. I rarely analyse the conversations I have had with others after they occur	☐	☐	☐	☐	☐	☐
12. When I meet a new person I usually evaluate how well he or she is doing on various dimensions (e.g. looks, achievements, social status, clothes)	☐	☐	☐	☐	☐	☐

6. Section 6 (of 8)

Read each of the following statements and place a tick in the appropriate box. Do not spend too much time on any statement. Answer quickly and honestly.

	Rarely/Never	*Occasionally*	*Often*	*Almost Always/Always*
1. I plan tasks carefully	☐	☐	☐	☐
2. I do things without thinking	☐	☐	☐	☐
3. I am happy-go-lucky	☐	☐	☐	☐
4. I have "racing" thoughts	☐	☐	☐	☐
5. I plan trips well ahead of time	☐	☐	☐	☐
6. I am self-controlled	☐	☐	☐	☐
7. I concentrate easily	☐	☐	☐	☐
8. I save regularly	☐	☐	☐	☐
9. I find it hard to sit still for long periods of time	☐	☐	☐	☐
10. I am a careful thinker	☐	☐	☐	☐
11. I plan for job security	☐	☐	☐	☐
12. I say things without thinking	☐	☐	☐	☐
13. I like to think about complex problems	☐	☐	☐	☐
14. I change jobs	☐	☐	☐	☐
15. I act "on impulse"	☐	☐	☐	☐
16. I get easily bored when solving thought problems	☐	☐	☐	☐
17. I have regular medical/dental check-ups	☐	☐	☐	☐
18. I act on the spur of the moment	☐	☐	☐	☐
19. I am a steady thinker	☐	☐	☐	☐
20. I change where I live	☐	☐	☐	☐
21. I buy things on impulse	☐	☐	☐	☐
22. I finish what I start	☐	☐	☐	☐
23. I walk and move fast	☐	☐	☐	☐
24. I solve problems by trial-and-error	☐	☐	☐	☐
25. I spend or charge more than I earn	☐	☐	☐	☐
26. I talk fast	☐	☐	☐	☐
27. I have outside thoughts when thinking	☐	☐	☐	☐

	Rarely/Never	*Occasionally*	*Often*	*Almost Always/Always*
28. I am more interested in the present than the future	☐	☐	☐	☐
29. I am restless at lectures or talks	☐	☐	☐	☐
30. I plan for the future	☐	☐	☐	☐

7. Section 7 (of 8)

Here are a number of characteristics that may or may not apply to you.
For example, do you agree that you are someone who likes to spend time with others?
Please respond according to the following scale to indicate how much you agree or disagree with each statement.

I see myself as someone who...

	Strongly Agree	*Somewhat Agree*	*Neither Agree or Disagree*	*Somewhat Disagree*	*Strongly Disagree*
1. **Does a thorough job**	☐	☐	☐	☐	☐
2. **Can be somewhat careless**	☐	☐	☐	☐	☐
3. **Is a reliable worker**	☐	☐	☐	☐	☐
4. **Tends to be disorganised**	☐	☐	☐	☐	☐
5. **Tends to be lazy**	☐	☐	☐	☐	☐
6. **Perseveres until the task is finished**	☐	☐	☐	☐	☐
7. **Does things efficiently**	☐	☐	☐	☐	☐
8. Makes plans and follows through with them	☐	☐	☐	☐	☐
9. Is easily distracted	☐	☐	☐	☐	☐

8. Section 8 (of 8)

Please read each of the following statements and decide how much you agree or disagree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	***Slightly Disagree***	***Neither Agree or Disagree***	***Slightly Agree***	***Strongly Agree***
1. I usually trust machines until there is a reason not to	☐	☐	☐	☐	☐
2. For the most part, I distrust machines	☐	☐	☐	☐	☐
3. In general, I would rely on a machine to assist me	☐	☐	☐	☐	☐
4. My tendency to trust machines is high	☐	☐	☐	☐	☐
5. It is easy for me to trust machines to do their job	☐	☐	☐	☐	☐
6. I am likely to trust a machine even when I have little knowledge about it	☐	☐	☐	☐	☐

Participant Debrief - please retain

Thank you.

You have now completed the questionnaire which signifies the end of your involvement with this study. The researchers would like to thank you for taking part.

This questionnaire was designed to explore the relationship between certain cognitive traits and propensity towards trusting autonomous tools and systems.

Can I contact the researcher for more information or to provide feedback about the study?

YES- we value your feedback and reflections on the research study to aid in future research. If you have any questions about the current study, would like to provide any feedback (positive or negative) or would like a copy of the results of the research then please contact the researchers at c.barrett-pink@liverpool.ac.uk

C Appendix Three

ARCS STUDY ADVERT, PARTICIPANT INFORMATION, CONSENT FORM AND DEBRIEF SHEET

1.1 STUDY ADVERT & FLYER

UNIVERSITY OF LIVERPOOL

[dstl]

PARTICIPANTS WANTED!

EXPLORING THE UPTAKE OF AN AUTOMATED DECISION SUPPORT SYSTEM

WHAT WILL YOU BE ASKED TO DO?

Session 1:

- Complete Scenario A using a simulated operations room.
- Complete a short series of questionnaires on your experiences.

The task will take no longer than 2 hours.

Session 2 (two weeks later):

- Complete scenario B using a simulated operations room.
- Complete a short series of questionnaires on your experiences

The task will take no longer than 2 hours.

Following completion of both sessions, you will receive £10 for your time!

HOW WILL YOUR PARTICIPATION HELP?

The results from this study will be used to further our understanding of the relationship between human and automated system.

WHAT DO I DO IF I AM INTERESTED IN PARTICIPATING?

If you would like to be involved in this study, and/or would like any more information on the researcher, what the study entails, please email

Chloë Barrett-Pink at c.barrett-pink@liverpool.ac.uk

WE LOOK FORWARD TO HEARING FROM YOU!

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

Automated Decision Support Study
c.barrett-pink@liverpool.ac.uk

EXPLORING THE UPTAKE OF AN AUTOMATED DECISION SUPPORT SYSTEM

WHAT WILL YOU BE ASKED TO DO?

Session 1
- Complete scenario A using a simulated operations room.
- Complete a short series of questionnaires on your experiences.

The task will take no longer than 2 hours.

Session 2 (two weeks later)
- Complete a scenario B using a simulated operations room.
- Complete a short series of questionnaires on your experiences.

The task will take no longer than 2 hours.

Following completion of both sessions you will receive £10 for your time.

WHAT DO I DO IF I AM INTERESTED IN PARTICIPATING?
If you would like to be involved in this study, and/or would like any more information on the researcher or what the study entails, please email Chloe Barrett-Pink at c.barrett-pink@liverpool.ac.uk

WE LOOK FORWARD TO HEARING FROM YOU!

UNIVERSITY OF LIVERPOOL

1.2 ARCS PARTICIPANT STUDY INFORMATION

Exploring the uptake of an automated decision support system

MoDREC Application No : 785/MoDREC/16

Invitation to Take Part

We would like to invite you to take part in a research study exploring the use of an automated system designed to aid individuals' decision making as part a PhD project funded by the Defence Science and Technology Laboratory's (Dstl) National PhD Scheme. 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 us if you would like more information or if there is anything that you do not understand.

What is the Purpose of the Research ?

The aim of this study is to explore the motivational and cognitive processes behind decision making during a classification task.

Who is Doing This Research ?

This research is being conducted by Miss Chloe Barrett-Pink as part of her PhD project supervised by Prof. Simon Maskell and Prof. Laurence Alison. The technical partner for this research is Mr Robert Walden. This research is funded by MOD via Dstl's National PhD Scheme.

Why Have I Been Invited to Take Part ?
You have been invited to take part as you replied to the advert posted around campus. This research aims to help further our understanding of the relationship between human and automated system.

Do I Have to Take Part ?
No, your participation is completely voluntary and you are free to withdraw from the research at any time should you decide you no longer wish to take part without giving a reason.

What Will I Be Asked to Do ?
This is a two-part study. You will be asked to complete a realistic Naval scenario involving the classification of icons on the radar picture into neutral, friendly or hostile aircraft.

It is important to be as accurate and confident as possible with each classification decision. First, a short training presentation will be given to provide you with information on the background to the task and how the system works. You will then complete a 15-minute training round to familiarise yourself with the task and the system. Following this, the main 30-minute task will begin. During which there may be the option to use an automated decision support system to help you complete the task. Following the main task, you will be asked to complete a series of questionnaires which should take no longer than 30 minutes. In total the whole task will take no longer than 2 hours.

Two weeks later you will be asked to return and complete the task again. A new scenario will be presented within which all icons on the radar must be classified as accurately and quickly as possible.

All results collected will be anonymised and will be used for this study and future research within this project. If you wish to remain updated with the findings of this study and the overall project that your participation will aid, please contact the researcher.

What are the Benefits of Taking Part ?
None to you directly, however you will receive £10 for your time in completing both parts of the experiment.

What are the Possible Disadvantages and Risks of Taking Part ?
There are no significant risks or disadvantages anticipated from taking part. However, during the training segment you will be exposed to a YouTube clip which may be potentially upsetting. You will be given prior warning of this material and provided with the option to not view this video clip.

Can I Withdraw from the Research and What Will Happen If I Withdraw ?
Yes, you are free to withdraw from the research at any time should you no longer wish to take part and you do not have to give a reason for doing so. Any data that has been collected up to the point at which you choose to withdraw will be removed from analysis and destroyed.

Are There Any Expenses and Payments Which I Will Get ?
Yes, upon completion of both parts of the experiment you will receive £10 for your time.

Whom Do I Contact If I Have Any Questions?

If you would like any more information on the researcher and what the study entails please email Prof. Simon Maskell (smaskell@liverpool.ac.uk) or Chloe Barrett-Pink (c.barrett-pink@liverpool.ac.uk).

Whom Do I Contact If I Have a Complaint?

In accordance with standard practice at the University of Liverpool, if you have any complaints please contact the University Ethics Committee (ethics@liverpool.ac.uk)

What Happens If I Suffer Any Harm ?

In the unlikely event that you suffer any harm as a direct result of taking part in this study, you can apply for compensation under the MoD's 'No-Fault Compensation Scheme.

Will My Records Be Kept Confidential ?

All the data collected will be anonymised and stored in a locked filing cabinet and/or on a password protected storage device during the course of this research. Following completion all consent forms will be forwarded to the MoDREC Secretariant to be stored for 50 years in accordance with extant UK legislation and MoD policy.

Who is Organising and Funding the Research ?

The research is part of a Dstl PhD funded studentship

Who Has Reviewed the Study ?

This study has been reviewed and given favourable opinion by the Ministry of Defence Research Ethics Committee (MoDREC).

This study has also been reviewed and given favourable opinion by the University of Liverpool Ethics Committee.

Further Information and Contact Details

Name : Miss Chloe Barrett-Pink
Address : Univeristy of Liverpool, School of Psychological Sciences, Bedford Street South, L69 7ZA
Tel No : +44(0)151 794 3936
E-mail : c.barrett-pink@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.

Thank you for taking the time to read this information sheet. If you would like to take part in this research then please read and sign the attached consent form.

[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).

1.3 Participant Consent form

PARTICIPANT CONSENT FORM

Study title: Exploring the uptake of an automated decision support system
Name of Researcher: Chloë Barrett-Pink

Please tick boxes

1. The nature, aims and risks of the research have been explained to me. I have read and understood the Information for Participants and understand what is expected of me. All my questions have been answered fully to my satisfaction. ☐

2. I understand that if I decide at any time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately without having to give a reason. I also understand that I may be withdrawn from it at any time, and that in neither case will this be held against me in subsequent dealings with the Ministry of Defence. ☐

3. I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998. ☐

4. I agree to volunteer as a participant for the study described in the information sheet and give full consent. ☐

Participant's Statement:

I ………………………………………………….

agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information for Participants about the project, and understand what the research study involves.
Signed: ___________________________ Date:___________

Investigator's Statement:

I Chloë Barrett-Pink confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the Participant

Signed: ______________________________ Date: ___________

Authorising Signatures

The information supplied above is to the best of my knowledge and belief accurate. I clearly understand my obligations and the rights of research participants, particularly concerning recruitment of participants and obtaining valid consent.

Signature of Chief Investigator

………………………………………………… **Date:**

Name and Contact Details of Chief Investigator:
Prof. Simon Maskell
Email: smaskell@liverpool.ac.uk
Telephone: +44(0)151 794 4573

Student Researcher:
Name: Chloë Barrett-Pink
Email: hlcbarre@liverpool.ac.uk

1.4 Participant debrief sheet

DEBRIEF AND FURTHER INFORMATION SHEET

Thank you

You have now completed the experiment which signifies the end of your involvement with this study. The researchers would like to thank you for taking part.

More information

The experiment was designed to explore the uptake and use of an autonomous system designed to support decision making within a complex environment.

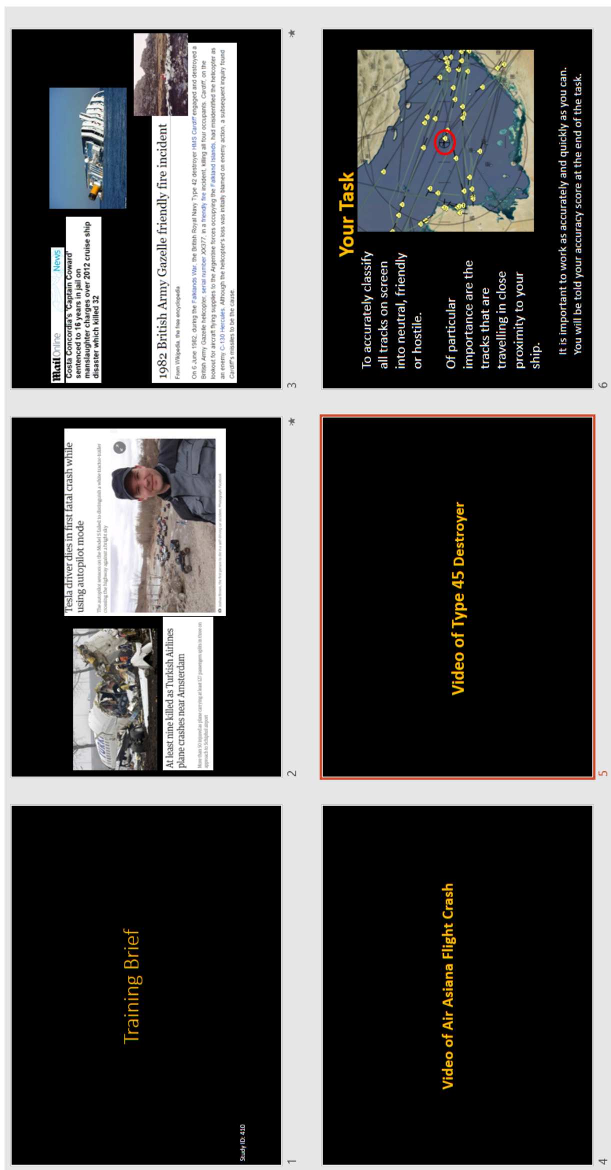
Can I contact the researchers for more information or to provide feedback about the study?

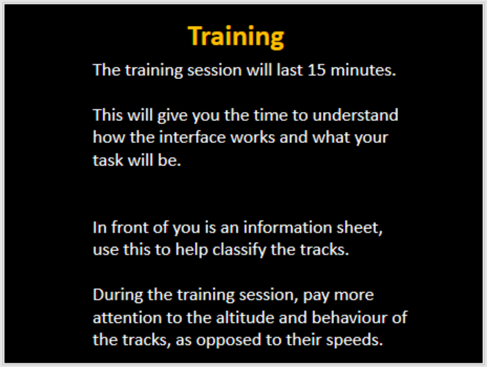
Yes – we value your feedback and reflections on the research study to aid future research. If you have any questions about the current study or would like to provide any feedback (positive or negative) then please contact the researcher at c.barrett-pink@liverpool.ac.uk

D Appendix Four

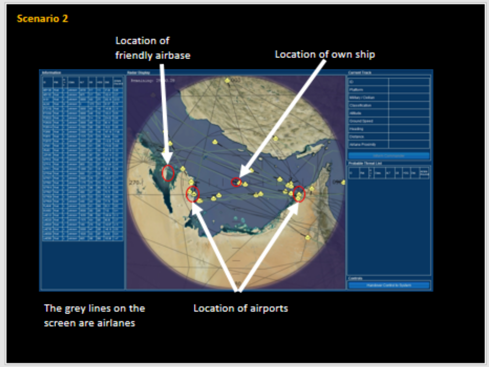
ARCS experiment information

1.1 Training Slides (The scenarios and information for each condition was separated when presented to participants)

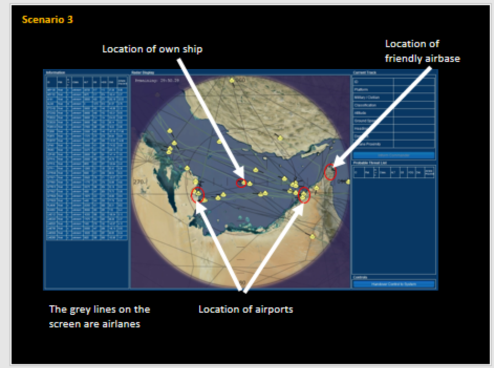



Training
The training session will last 15 minutes.
This will give you the time to understand how the interface works and what your task will be.
In front of you is an information sheet, use this to help classify the tracks.
During the training session, pay more attention to the altitude and behaviour of the tracks, as opposed to their speeds.

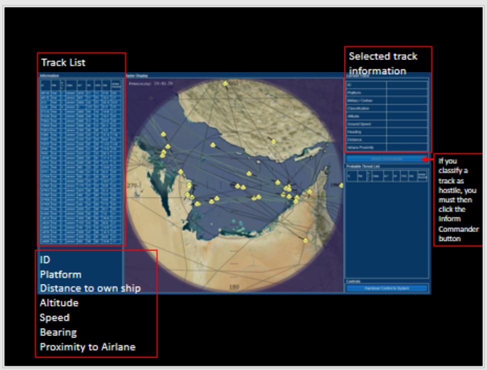
7


Scenario 2
Location of friendly airbase
Location of own ship
The grey lines on the screen are airlanes
Location of airports

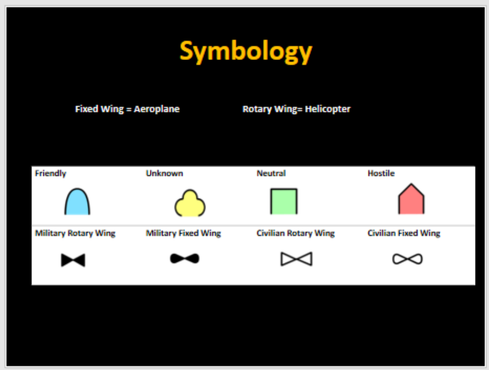
8


Scenario 3
Location of own ship
Location of friendly airbase
The grey lines on the screen are airlanes
Location of airports

9


Track List
Selected track information
If you classify a track as hostile, you must then click the Inform Commander button
ID
Platform
Distance to own ship
Altitude
Speed
Bearing
Proximity to Airlane

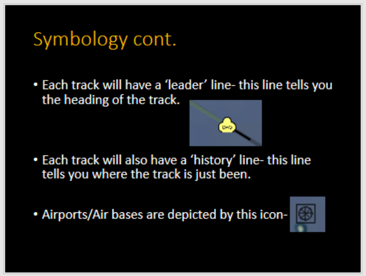
10


Symbology
Fixed Wing = Aeroplane
Rotary Wing= Helicopter
Friendly
Unknown
Neutral
Hostile
Military Rotary Wing
Military Fixed Wing
Civilian Rotary Wing
Civilian Fixed Wing

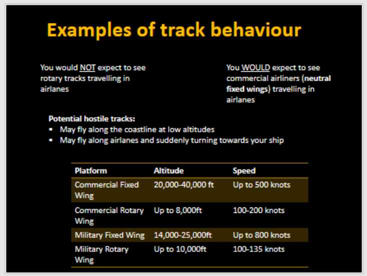
11


More examples...
Friendly Military Fixed Wing
Neutral Civilian Rotary Wing
Hostile Military Fixed Wing
Unknown Civilian Rotary Wing
Friendly Military Rotary Wing
Neutral Civilian Fixed Wing
Hostile Military Rotary Wing
Unknown Civilian Fixed Wing
Neutral Military Rotary Wing
Unknown Military Rotary Wing
Neutral Military Fixed Wing
Unknown Military Fixed Wing

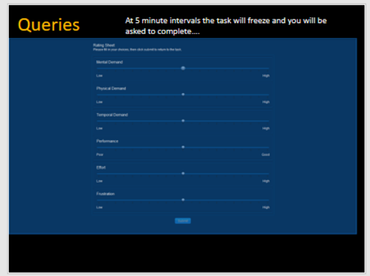
12



13

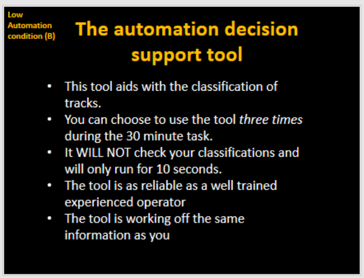


14

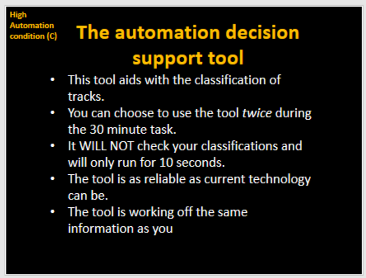


15

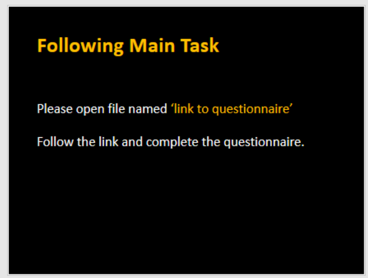
16



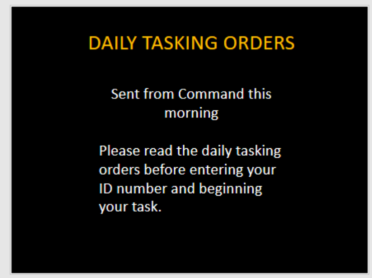
17



18



19



20

1.2 INFORMATION SHEET

INFORMATION SHEET

Platform	Altitude	Speed
Commercial Fixed Wing	20,000-40,000ft	Up to 500knots
Commercial Rotary Wing	Up to 8,000ft	100-200knots
Military Fixed Wing	14,000-25,000ft	Up to 800knots
Military Rotary Wing	Up to 10,000ft	100-135knots

If information relating to a track meets these criteria you can assume the relevant platform.

Please be as accurate and confident with your decisions as possible.

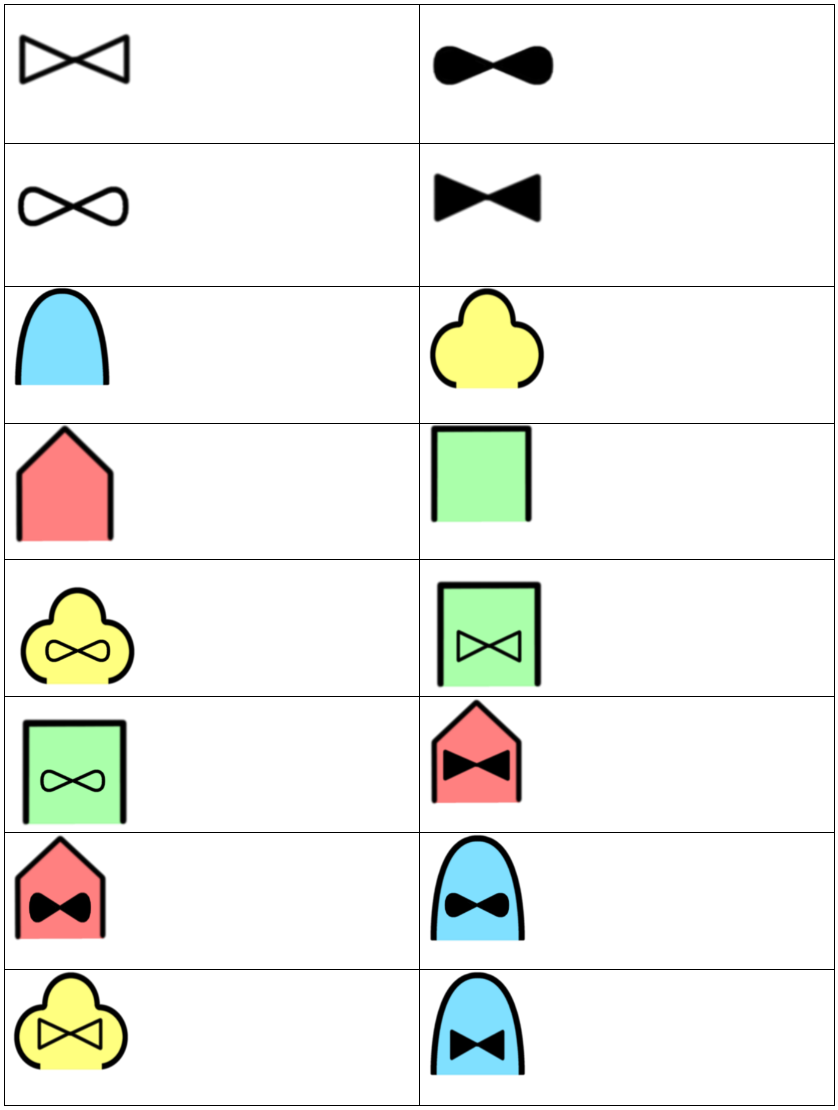
Note that the altitude and speed of each track is variable, requiring consistent monitoring to detect any changes.

Over time you will also notice behavioral patterns of each track.

For example:

- A Civilian or Military Rotary Wing *would not be* expected to travel within an airlane.
- A Commercial Fixed Wing *would be* expected to travel within an airlane.

1.3 Track Symbology



1.4 DEFINITIONS OF NASA-TLX WORKLOAD DIMENSIONS

MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard do you have to work (mentally or physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

E Appendix Five

QUESTIONNAIRE GIVEN TO PARTICIPANTS FOLLOWING ARCS

SECTION 1 (OF 8)

Please provide the following information about yourself by ticking the appropriate box or write your answer in the space provided.

I.D. Number: ………………

Gender: …………………………

What is your age in years? ………………………………………………………………

What course did you previously study? ……………………………………………………………………

What course are you currently studying? ………………………………………………………………….

SECTION 2 (OF 8) **FOR PARTICIPANTS IN CONDITION A**

1. From 1-10 how would you rate your performance (1=poor, 10=perfect) ……………………………..
2. On a scale of 1-10 how accountable did you feel towards the decisions you made? (1=not accountable, 10= very accountable) …………………………………
3. Did you recheck tracks once you made a classification decision? YES/NO
4. If YES, how often did you recheck and why?
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
5. If you could perform the scenario again, is there anything you would do differently?
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………
   ………………………………………………………………………………………………………

SECTION 2 (OF 8) **FOR PARTICIPANTS IN CONDITIONS B AND C**

1. From 1-10 how would you rate your performance (1=poor, 10=perfect) ……………………………..
2. On a scale of 1-10 how accountable did you feel towards the decisions you made? (1=not accountable, 10= very accountable) …………………………………
3. Did you choose the use the decision support system at any point(s) during the scenario? YES/NO
    a. Why did you choose to use the decision support system?……………………………………………………………………………………
    ………
    ………
    ………
    ………
    ………………………………………………
    ………
    ………
    ………
    ………
    ………………………………………………
4. Reflecting on your performance, were there any occasions that you think you should have used the decision support system but did not and why?
………
………
………
………
………
………
…………………………………………………………………
5. Reflecting back on the scenario, were there any occasions that you felt you could have used the decision support system earlier but did not?
………
………
………
………
………
………
………
………
……………………………………………………………………………………
6. What made you not choose the system at an earlier point?……………………………………………………………………………………
………
………
………
…………………………………………
………
………
………
………

....
....

7. Did you recheck tracks once you made a classification decision? YES/NO
8. If YES, how often did you recheck and why?
....
....
....
....
....
....
....
....
....
....
....
9. If you could perform the scenario again, is there anything you would do differently?
....
....
....
....
....
....
....
....
....
....
....
....

SECTION 3 (OF 8)

Please use the scale below to indicate the extent to which you agree or disagree with the following statements:

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Neutral*	*Moderately Agree*	*Agree*	*Strongly Agree*
1. I am good at "sizing up" situations	☐	☐	☐	☐	☐	☐	☐
2. I have a hard time making decisions when faced with difficult situations	☐	☐	☐	☐	☐	☐	☐
3. I consider multiple options before making a decision	☐	☐	☐	☐	☐	☐	☐
4. When I encounter difficult situations, I feel like I am losing control	☐	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Neutral*	*Moderately Agree*	*Agree*	*Strongly Agree*
5. I like to look at difficult situations from many different angles	☐	☐	☐	☐	☐	☐	☐
6. I seek additional information not immediately available before attributing causes to behaviour	☐	☐	☐	☐	☐	☐	☐
7. When encountering difficult situations, I become so stressed that I cannot think of a way to resolve the situation	☐	☐	☐	☐	☐	☐	☐
8. I try to think about things from another person's point of view	☐	☐	☐	☐	☐	☐	☐
9. I find it troublesome that there are so many different ways to deal with difficult situations	☐	☐	☐	☐	☐	☐	☐
10. I am good at putting myself in others' shoes	☐	☐	☐	☐	☐	☐	☐
11. When I encounter difficult situations, I just don't know what to do	☐	☐	☐	☐	☐	☐	☐
12. It is important to look at difficult situations from many angles	☐	☐	☐	☐	☐	☐	☐
13. When in difficult situations, I consider multiple options before deciding how to behave	☐	☐	☐	☐	☐	☐	☐
14. I often look at a situation from different view points	☐	☐	☐	☐	☐	☐	☐
15. I am capable of overcoming the	☐	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Neutral*	*Moderately Agree*	*Agree*	*Strongly Agree*
difficulties in life that I face							
16. I consider all the available facts and information when attributing causes to behaviour	☐	☐	☐	☐	☐	☐	☐
17. I feel I have no power to change things in difficult situations	☐	☐	☐	☐	☐	☐	☐
18. When I encounter difficult situations, I stop and try to think of several ways to resolve it	☐	☐	☐	☐	☐	☐	☐
19. I can think of more than one way to resolve a difficult situation I'm confronted with	☐	☐	☐	☐	☐	☐	☐
20. I consider multiple options before responding to difficult situations	☐	☐	☐	☐	☐	☐	☐

SECTION 4 (OF 8)

Read each of the following statements and decide how much you agree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
1. I think that having clear rules and order at work is essential for success	☐	☐	☐	☐	☐	☐
2. Even after I've made up my mind about something, I am always eager to consider a different option	☐	☐	☐	☐	☐	☐
3. I don't like situations that are uncertain	☐	☐	☐	☐	☐	☐
4. I dislike questions which could be answered in many different ways	☐	☐	☐	☐	☐	☐
5. I like to have friends who are unpredictable	☐	☐	☐	☐	☐	☐
6. I find that a well ordered life with regular hours suits my temperament	☐	☐	☐	☐	☐	☐
7. I enjoy the uncertainty of going into a new situation without knowing what might happen	☐	☐	☐	☐	☐	☐
8. When dining out, I like to go to places where I have been before so that I know what to expect	☐	☐	☐	☐	☐	☐
9. I feel uncomfortable when I don't understand the reason why an event occurred in my life	☐	☐	☐	☐	☐	☐

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
10. I feel irritated when one person disagrees with what everyone else in a group believes	☐	☐	☐	☐	☐	☐
11. I hate to change plans at the last minute	☐	☐	☐	☐	☐	☐
12. I would describe myself as indecisive	☐	☐	☐	☐	☐	☐
13. When I go shopping, I have difficulty deciding exactly what it is I want	☐	☐	☐	☐	☐	☐
14. When faced with a problem I usually see the one best solution very quickly	☐	☐	☐	☐	☐	☐
15. When I am confused about an important issue, I feel very upset	☐	☐	☐	☐	☐	☐
16. I tend to put off making important decisions until the last possible moment	☐	☐	☐	☐	☐	☐
17. I usually make important decisions quickly and confidently	☐	☐	☐	☐	☐	☐
18. I have never been late for an appointment or work	☐	☐	☐	☐	☐	☐
19. I think it is fun to change my plans at the last minute	☐	☐	☐	☐	☐	☐
20. My personal space is usually messy and disorganised	☐	☐	☐	☐	☐	☐
21. In most social conflicts, I can easily see which side is right and which is wrong	☐	☐	☐	☐	☐	☐

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
22. I have never known someone I did not like	☐	☐	☐	☐	☐	☐
23. I tend to struggle with most decisions	☐	☐	☐	☐	☐	☐
24. I believe orderliness and organisation are among the most important characteristics of a good student	☐	☐	☐	☐	☐	☐
25. When considering most conflict situations, I can usually see how both sides could be right	☐	☐	☐	☐	☐	☐
26. I don't like to be with people who are capable of unexpected actions	☐	☐	☐	☐	☐	☐
27. I prefer to socialise with familiar friends because I know what to expect from them	☐	☐	☐	☐	☐	☐
28. I think that I would learn best in a class that lacks clearly stated objectives and requirement	☐	☐	☐	☐	☐	☐
29. When thinking about a problem, I consider as many different opinions on the issue as possible	☐	☐	☐	☐	☐	☐
30. I don't like to go into a situation without knowing what I can expect from it	☐	☐	☐	☐	☐	☐
31. I like to know what people are thinking all the time	☐	☐	☐	☐	☐	☐

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
32. I dislike it when a person's statement could mean many different things	☐	☐	☐	☐	☐	☐
33. It's annoying to listen to someone who cannot seem to make up his or her mind	☐	☐	☐	☐	☐	☐
34. I find that establishing a consistent routine enables me to enjoy life more	☐	☐	☐	☐	☐	☐
35. I enjoy having a clear and structured mode of life	☐	☐	☐	☐	☐	☐
36. I prefer interacting with people whose opinions are very different from my own	☐	☐	☐	☐	☐	☐
37. I like to have a plan for everything and a place for everything	☐	☐	☐	☐	☐	☐
38. I feel uncomfortable when someone's meaning or intention is unclear to me	☐	☐	☐	☐	☐	☐
39. I believe that one should never engage in leisure activities	☐	☐	☐	☐	☐	☐
40. When trying to solve a problem I often see so many possible options that it's confusing	☐	☐	☐	☐	☐	☐
41. I always see many possible solutions to problems I face	☐	☐	☐	☐	☐	☐
42. I'd rather know bad news than stay in a state of uncertainty	☐	☐	☐	☐	☐	☐
43. I feel that there is no such thing as an honest mistake	☐	☐	☐	☐	☐	☐

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
44. I do not usually consult many different options before forming my own view	☐	☐	☐	☐	☐	☐
45. I dislike unpredictable situations	☐	☐	☐	☐	☐	☐
46. I have never hurt another person's feelings	☐	☐	☐	☐	☐	☐
47. I dislike the routine aspects of my work (studies)	☐	☐	☐	☐	☐	☐

SECTION 5.1 (OF 8)

Read each of the following statements and decide how much you agree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Moderately Agree*	*Agree*	*Strongly Agree*
1. I don't mind doing things even if they involve extra effort	☐	☐	☐	☐	☐	☐
2. I am a "workaholic"	☐	☐	☐	☐	☐	☐
3. I feel excited just before I am about to reach a goal	☐	☐	☐	☐	☐	☐
4. I enjoy actively doing things, more than just watching and observing	☐	☐	☐	☐	☐	☐
5. I am a "doer"	☐	☐	☒	☐	☐	☐
6. When I finish one project, I often wait awhile before getting started on a new one	☐	☐	☐	☐	☐	☐
7. When I decide to do something, I can't wait to get started	☐	☐	☐	☐	☐	☐
8. By the time I accomplish a task, I already have the next one in mind	☐	☐	☐	☐	☐	☐
9. I am a "low energy" person	☐	☐	☐	☐	☐	☐
10. Most of the time my thoughts are occupied with the task I wish to accomplish	☐	☐	☐	☐	☐	☐

	Strongly Disagree	Disagree	Slightly Disagree	Moderately Agree	Agree	Strongly Agree
11. When I get started on something, I usually persevere until I finish it	☐	☐	☐	☐	☐	☐
12. I am a "go-getter"	☐	☐	☐	☐	☐	☐

SECTION 5.2 (OF 8)

Read each of the following statements and decide how much you agree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	Disagree	Slightly Disagree	Moderately Agree	Agree	Strongly Agree
1. I never evaluate my social interactions with others after they occur	☐	☐	☐	☐	☐	☐
2. I spend a great deal of time taking inventory of my positive and negative characteristics	☐	☐	☐	☐	☐	☐
3. I like evaluating other people's plans	☐	☐	☐	☐	☐	☐
4. I often compare myself with other people	☐	☐	☐	☐	☐	☐
5. I don't spend much time thinking about ways others could improve themselves	☐	☐	☐	☐	☐	☐
6. I often critique work done by myself or others	☐	☐	☐	☐	☐	☐
7. I often feel that I am being evaluated by others	☐	☐	☐	☐	☐	☐
8. I am a critical person	☐	☐	☐	☐	☐	☐
9. I am very self-critical and self-conscious about what I am saying	☐	☐	☐	☐	☐	☐
10. I often think that other people's choices and decisions are wrong	☐	☐	☐	☐	☐	☐
11. I rarely analyse the conversations I have had with others after they occur	☐	☐	☐	☐	☐	☐
12. When I meet a new person I usually evaluate how well he or she is doing on various dimensions (e.g. looks, achievements, social status, clothes)	☐	☐	☐	☐	☐	☐

SECTION 6 (OF 8)

Read each of the following statements and place a tick in the appropriate box. Do not spend too much time on any statement. Answer quickly and honestly.

	Rarely/Never	*Occasionally*	*Often*	*Almost Always/Always*
1. I plan tasks carefully	☐	☐	☐	☐
2. I do things without thinking	☐	☐	☐	☐
3. I am happy-go-lucky	☐	☐	☐	☐
4. I have "racing" thoughts	☐	☐	☐	☐
5. I plan trips well ahead of time	☐	☐	☐	☐
6. I am self-controlled	☐	☐	☐	☐
7. I concentrate easily	☐	☐	☐	☐
8. I save regularly	☐	☐	☐	☐
9. I find it hard to sit still for long periods of time	☐	☐	☐	☐
10. I am a careful thinker	☐	☐	☐	☐
11. I plan for job security	☐	☐	☐	☐
12. I say things without thinking	☐	☐	☐	☐
13. I like to think about complex problems	☐	☐	☐	☐
14. I change jobs	☐	☐	☐	☐
15. I act "on impulse"	☐	☐	☐	☐
16. I get easily bored when solving thought problems	☐	☐	☐	☐
17. I have regular medical/dental check-ups	☐	☐	☐	☐
18. I act on the spur of the moment	☐	☐	☐	☐
19. I am a steady thinker	☐	☐	☐	☐
20. I change where I live	☐	☐	☐	☐
21. I buy things on impulse	☐	☐	☐	☐
22. I finish what I start	☐	☐	☐	☐
23. I walk and move fast	☐	☐	☐	☐
24. I solve problems by trial-and-error	☐	☐	☐	☐
25. I spend or charge more than I earn	☐	☐	☐	☐
26. I talk fast	☐	☐	☐	☐
27. I have outside thoughts when thinking	☐	☐	☐	☐
28. I am more interested in the present than the future	☐	☐	☐	☐

	Rarely/Never	*Occasionally*	*Often*	*Almost Always/Always*
29. I am restless at lectures or talks	☐	☐	☐	☐
30. I plan for the future	☐	☐	☐	☐

SECTION 7 (OF 8)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please respond according to the following scale to indicate how much you agree or disagree with each statement.

I see myself as someone who…		*Strongly Disagree*	*Disagree a little*	*Neither Agree or Disagree*	*Agree a little*	*Strongly Agree*
	1. Does a thorough job					
	2. Can be somewhat careless					
	3. Is a reliable worker					
	4. Tends to be disorganised					
	5. Tends to be lazy					
	6. Perseveres until the task is finished					
	7. Does things efficiently					
	8. Makes plans and follows through with them					
	9. Is easily distracted					

SECTION 8 (OF 8)

Please read each of the following statements and decide how much you agree or disagree with each according to your beliefs and experiences. Please respond according to the following scale:

	Strongly Disagree	*Slightly Disagree*	*Neither Agree or Disagree*	*Slightly Agree*	*Strongly Agree*
1. I usually trust machines until there is a reason not to	☐	☐	☐	☐	☐
2. For the most part, I distrust machines	☐	☐	☐	☐	☐
3. In general, I would rely on a machine to assist me	☐	☐	☐	☐	☐
4. My tendency to trust machines is high	☐	☐	☐	☐	☐
5. It is easy for me to trust machines to do their job	☐	☐	☐	☐	☐
6. I am likely to trust a machine even when I have little knowledge about it	☐	☐	☐	☐	☐

F Appendix Six

Accountability Primes

1.1 Primes for Condition A – Control group

DAILY TASKING ORDERS

MISSION

Your mission is to build the air picture through classifying each track on the radar. Of particular importance is to identify any possible hostile targets and inform Command of their presence.

Your overall performance is judged based on your performance individually. It is vital that the task is performed as accurately as possible, mistakes and errors are not acceptable.

<u>You will be held individually accountable to Command and responsible for any errors made.</u>

You will be expected to provide an account of your performance.

DAILY TASKING ORDERS

MISSION

Your mission is to build the air picture through classifying each track on the radar. Of particular importance is to identify any possible hostile targets and inform Command of their presence.

Your overall performance is judged based on your performance individually. Please try to be as accurate as possible, however, as this is a complex task it is expected that small errors may be made.

<u>You will not be held individually accountable to Command for any errors made.</u>

You will be expected to provide an account of your performance.

1.2 PRIMES FOR CONDITIONS B & C

DAILY TASKING ORDERS

MISSION

Your mission is to build the air picture through classifying each track on the radar. Of particular importance is to identify any possible hostile targets and inform Command of their presence.

Your overall performance is judged based on your performance individually and as a team with the tool. It is vital that the task is performed as accurately as possible, mistakes and errors are not acceptable.

You will be held individually accountable to Command and responsible for any errors made.

You will be expected to provide an account of your performance.

DAILY TASKING ORDERS

MISSION

Your mission is to build the air picture through classifying each track on the radar. Of particular importance is to identify any possible hostile targets and inform Command of their presence.

Your overall performance is judged based on your performance individually and as a team with the tool. Please try to be as accurate as possible, however, as this is a complex task it is expected that small errors may be made.

You will not be held individually accountable to Command for any errors made.

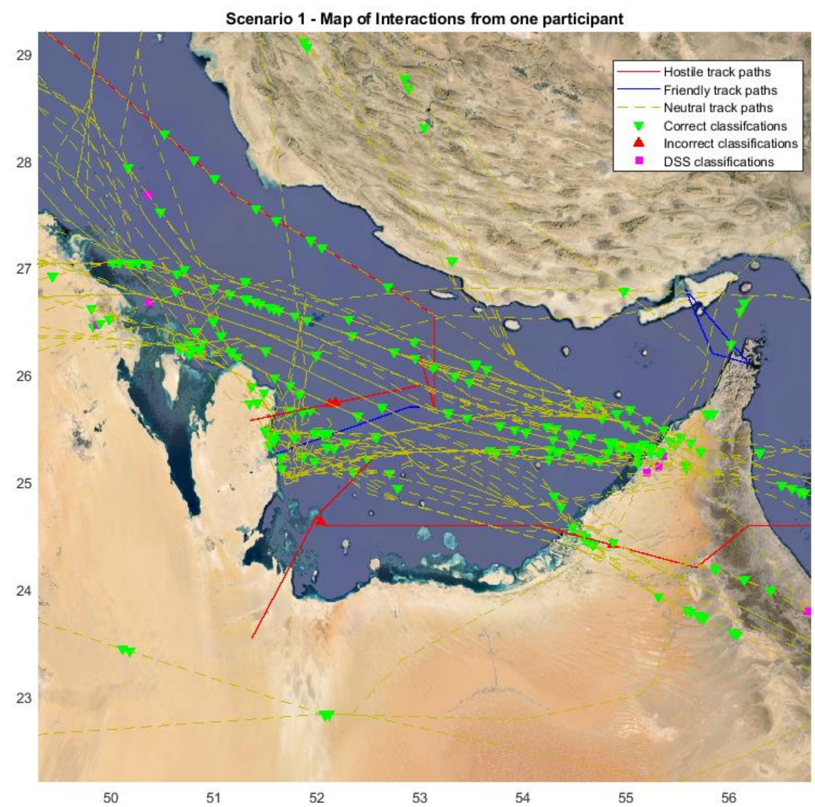
You will be expected to provide an account of your performance.

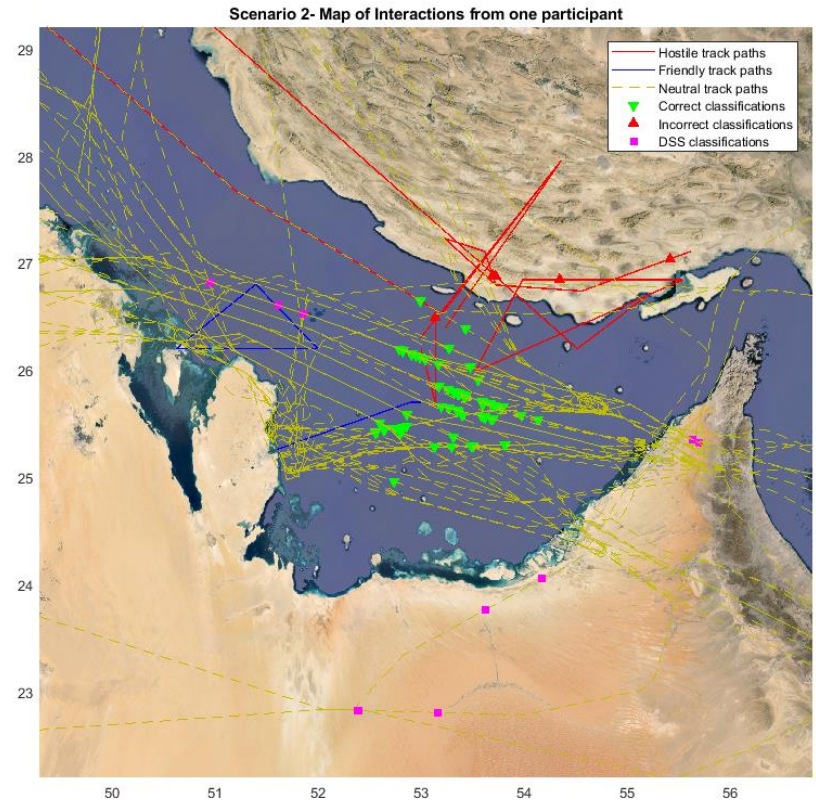
G Appendix Seven

EXAMPLES OF ALTERNATIVE WAYS TO EXPLORE THE DATA COLLECTED FROM ARCS

1.1 PARTICIPANT DECISION MAKING

Using MATLAB it is possible to generate an individual map of interaction for each participant. These detail the flight paths of each track, the tracks participants correctly and incorrectly identified, as well as the tracks that the automated system classified (see Figures below).





With the data that ARCS collects it is also possible to visually playback each completed scenario. From this it is possible to explore if certain search strategies equate to more accurate decision making. Although this thesis did not specifically look at search strategies ARCS has the potential to provide this information if future research required it.