The Role of Experience in the Susceptibility to Confirmation Bias in Pilots

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Arts

in

Psychology

at Massey University, Albany,

New Zealand.

Jaime Rowntree

2012

Abstract

Confirmation bias refers to the tendency of an individual to prioritise and seek out evidence that confirms their theory or hypothesis and avoid or place little importance on disconfirming information. In the field of aviation, confirmation bias can have disastrous consequences and has been implicated in several aviation disasters. Despite the potentially fatal consequences, little research has systematically explored the underlying causes of confirmation bias in pilots. The following research examined the role of experience in the susceptibility to confirmation bias in pilots utilising an aviation themed location discovery task. To assess the relationship between flying experience and susceptibility to confirmation bias, 53 participants (23 non-pilots with no prior flying experience, 13 novice pilots with between 0-200 hours of logged flight time, and 17 experienced pilots with between 220-15000 hours of logged flight time) were asked to complete an online map-based location discovery task, which required participants to imagine that they were unsure of their location in four aviation themed scenarios. They then had to select, out of three features given, which feature would be the most useful for helping them to decide on their current location. Two out of the three features provided incorrect confirming (positive) tests of the pilot's hypothesised location and one feature provided the correct disconfirming (negative) test of their hypothesised location. Results indicated that overall, participants primarily utilised a hypothesis-confirming strategy on the task. No relationship between experience and a participant's susceptibility to confirmation bias was identified. A thematic analysis of the comments provided by participants was completed, which illustrated that participants were fairly consistent in the decision-making strategy that they used when reasoning about their location across each of the four scenarios. Interestingly, non-pilots and novice pilots primarily utilised a hypothesis-confirming approach most regularly in their feature selection. By contrast, it appears that the experienced pilot group primarily utilised a strategy that favoured the selection of manufactured objects and large objects. Future research should focus on discovering the mechanisms underlying confirmation bias and the identification of groups of people who are less susceptible to it. This information can then be used to create a model of confirmation bias outlining interventions that can be used to reduce or eliminate its effects.

Acknowledgements

I would like to express my deepest thanks to my husband Daniel. Your support, encouragement, and patience have been incredible.

Thank you very much to my supervisors Dr. Stephen Hill and Dr. Andrew Gilbey. Your wise words, sage advice, patience, and support have been vital to the completion of my research.

A very big thank you to all of my family, friends, and work colleagues. Thank you for your support, inspiration, kind words, and laughter, which have kept me going.

Table of Contents

Chapter **Page Number**

List of Tables

List of Figures

List of Appendices

Chapter One

Confirmation Bias in Hypothesis Testing

1.1 Chapter Introduction

The occurrence of human errors and cognitive biases are inevitable and having an understanding of the human factors underlying these human errors and cognitive biases is useful to peoples' understanding of the decision-making processes that people engage in when solving problems. When faced with a problem, people often have to assess and evaluate alternative hypotheses to decide on the correct course of action. Prior research into hypothesis-testing strategies has illustrated a distinct tendency for participants to follow a hypothesis-confirming approach to their hypothesis testing and evaluation. This bias is referred to as *confirmation bias* and it has been illustrated repeatedly in research studies over the last few decades since P.C. Wason's original research displaying the phenomenon in 1960. The following chapter looks at the importance and role of schemas in the hypothesis-testing process, explores Wason's original rule discovery task and the limitations of his research. The chapter also looks at alternative explanations of confirmation bias and participants' hypothesis-testing strategies. These include Klayman and Ha's (1987) positive testing strategy and Farris and Revlin's (1989a, 1989b) counterfactual strategy.

1.2 Schemas and Knowledge Patterns

Schemas are defined as organised patterns of knowledge that have formed as a result of a persons continued exposure, interaction, or experience with an object or situation (Chen & Mo, 2004). They serve to instruct people on how to react in a given situation based on their accumulated knowledge of how familiar situations normally play out (Klayman, 1995). Consequently, the decision-making strategies that people utilise and the solutions they choose to problems, are primarily based on peoples' schemas and knowledge patterns that have formed over time (Bilalic, McLeod, & Gobet, 2010; Chen & Mo, 2004; Hooey & Foyle, 2006). Predominantly, schemas are advantageous as they allow people to quickly access prior knowledge on how to react to a problem or situation. They also allow people to add to their base of evidence on the most efficient and effective actions to take in any given scenario. As soon as an individual recognises a familiar problem or situation, existing schemas relevant to the task are activated, allowing people to draw from this knowledge to reach a conclusion. Individuals come to rely heavily on their set of schemas but they are narrowly limited to an individual's previous experiences and knowledge of certain events and situations (Fox & Rottenstreich, 2003). More experienced individuals therefore have a larger range of schemas to draw from in any given scenario (Adams, 1993; Bilalic et al., 2010; O'Hare, Mullen, Wiggins, & Molesworth, 2008).

Schemas encourage individuals to focus on information and evidence presented within a situation that is relevant to the schemas each person has available and disregard information that does not correspond with their existing schemas. As a result, people have a set of expectancies about the way in which certain situations will unfold (Bruner & Postman, 1949). Over time, these expectancies are reinforced by the outcome of life events and they continue to be cemented into schemas. As long as the schemas continue to be reinforced, they become self-fulfilling, in the sense that once a schema has been constructed, ambiguous evidence tends to be interpreted in a way that reinforces the explanatory framework given by the schema (Bilalic et al., 2010). Bruner and Postman (1949) investigated the pervasiveness of peoples' schemas, by assessing their reactions to information that contradicted previously held knowledge. Participants were presented with five playing cards where between one and four of the cards were considered incongruent with existing knowledge, as the suit and colour had been reversed. For example, two of the incongruent cards that were utilised were a red two of spades and a black three of hearts. Cards were briefly shown to participants at timed intervals and the experiment only moved on to the next card when the current card had been correctly recognised, or the time limit for that card had been reached. The average length of time required to correctly identify a normal card was 28 milliseconds and the average length of time required to correctly identify an incongruent card was significantly higher at 114 milliseconds (Bruner & Postman, 1949). Approximately 96% of participants attempted to conform the incongruent playing cards to their existing schema on how playing cards should be constructed, by either altering the colour or suit of the erroneous card to correspond with their existing knowledge (Bruner & Postman, 1949). Bruner and Postman (1949) argue that the longer length of time taken for participants to identify incongruous playing cards, was a result of resistance being experienced by the participants to information being received that was in direct opposition to their pre-existing knowledge.

Bruner and Postman (1949) concluded that the organisation of schemas is heavily reliant on a person's individual interactions with situations and items in their environment, which aids in the formation and reinforcement of schemas over time. When their expectations of phenomena are disrupted, people display a strong resistance to recognising any incongruent information. Theorising concerning such resistance to hypothesis disconfirming information was not significantly progressed until Wason (1960) created the rule discovery task. His research confirmed the existence of a pervasive cognitive bias that is characterised by peoples' primary tendency to search for confirmatory information in hypothesis testing and evaluation tasks.

1.3 Confirmation Bias in the Hypothesis Testing Tradition

> Four centuries ago, Francis Bacon pointed out the unfortunate human tendency to ignore new evidence that could undermine a firmly held opinion (Bacon, 1620/1939, p. 36, quoted by Bilalic et al., 2010, p. 111).

When faced with problems across a range of situations, people have to evaluate possible hypotheses to determine the cause of the current state of affairs. Individuals then perform tests of their possible hypotheses to produce evidence to help them decide on the correct course of action to solve the current problem (Tschirgi, 1980). Throughout this hypothesis testing process, people rely on their schemas and patterns of prior knowledge to reach a satisfactory conclusion to problems quickly and efficiently. As most real world situations are complex and contain a large number of features and potential evidence to find, assess, and evaluate, peoples' schemas play a vital role in directing individuals on where to focus their attention. Often it will not be possible for individuals to systematically exhaust all of the possible alternative hypotheses or conclusively be able to state that all evidence and alternative solutions were found and reviewed (Nickerson, 1996). Even if a full examination of the entire universe of possible evidence was plausible, one still could not prove that a hypothesised rule (R_H) is true, regardless of the amount of testing and evaluation that had been carried out. A hypothesis cannot be proved correct by relying only on confirmatory evidence regardless of the amount of evidence received, but a single observation of a piece of disconfirming evidence will be enough to quash the belief (Silverman, 1992). As it only takes one piece of evidence to disconfirm a theory, it would seem more plausible to search for disconfirming evidence rather than confirming evidence during the hypothesis testing process (Wason, 1960).

Despite this, the hypothesis testing literature shows that participants repeatedly display a robust and persistent bias towards information gained in the hypothesis testing process that confirms their pre-existing schemas (Nickerson, 1998). This inherent bias has been labelled as a *confirmation bias* and it is characterised as the tendency of an individual to prioritise and seek out evidence that confirms their theory or hypothesis and avoid or place little importance on disconfirming information (Nickerson, 1998; Rassin, 2008). The concept of confirmation bias originated from early research conducted by P.C. Wason in the 1960's (Wason, 1960, 1968) with the rule discovery task (RDT) experiment[1].

Wason's works were based on his own background knowledge and personal interests into the research of science philosopher Karl Popper, who posited that the optimum way to evaluate a hypothesis was to attempt to locate or source evidence to disprove it or to engage in a conscious effort to falsify it (Austerweil & Griffiths, 2011). Wason believed that "individuals are biased through a long learning process to seek and expect a simple correspondence to hold between sentences and states of affairs" (Wason, 1968, p. 280). By the time an individual reaches adulthood, these cognitive processes are so ingrained that it is difficult for an individual to adopt any other method of hypothesis evaluation. Often, a hypothesis-confirming evaluation method is successful in gaining the results that the individual is looking for and therefore this method of hypothesis testing becomes the dominant process, despite it not always being optimal.

Wason's original RDT experiment was first carried out in 1960, with the aim of illustrating that the majority of people do not evaluate and test hypotheses in the most effective way. It involved participants being given a set of three numbers, e.g., 2-4-6, that they were told conformed to a rule that the researcher had in mind. Participants were then instructed to try and work out this rule by suggesting further sets of triples to the researcher, who provided feedback on whether the proposed triple conformed to the rule

[1] The term confirmation bias was not coined by Wason himself. It was not until research conducted by Mynatt, Doherty, and Tweney (1977) that the term was used in this context.

or not. Despite the experimenter's original rule being simple in nature, described as any triple containing an ascending series of numbers, participants usually provided 'narrower' potential hypotheses, e.g., even numbers, increasing by two (Wason, 1960). These salient features of the original triple drew the participant's focus and often prompted generations of further triples to test the rule following this assumption, such as 8-10-12. As this triple is also consistent with the experimenter's rule of 'any ascending numbers', the participant received positive feedback which provided further support for their initial hypothesis. Unfortunately, in situations where the evidence received is consistent with a range of alternative hypotheses and the true rule (R_T), any positive feedback received is not diagnostically useful, as it does not give any feedback on the actual distinct properties of the true rule (Wason, 1960). Positive feedback in this situation can lead participants to erroneously believe that their current hypothesis is correct.

When participants felt confident that they had discovered the rule, they were asked to stop generating triples and announce their rule to the researcher. Overall, participants performed poorly on the RDT, with only 21% of participants reaching the correct conclusion on their first attempt and over 30% of participants providing two or more incorrect conclusions (Wason, 1960). Overwhelmingly, participants primarily sought out information to confirm their hypothesis and Wason (1960) proposed that participants would have had more success on the RDT if they had tested triples that did not conform to their original hypothesis. For example, if a participant's original hypothesis was even numbers ascending by two, the best hypothesis to suggest would be one that did not conform to this hypothesis, such as the number triple 3-9-17. This number triple would receive a positive response from the researcher, as it conforms to the true rule of ascending numbers and at the same time, it would allow the participant to correctly discount the hypothesis of even numbers, ascending by two. This strategy is postulated to be more diagnostic than a strategy that seeks primarily hypothesis-confirming information (Wason, 1960).

Wason (1960) believed that the failure of participants to come to the correct conclusion was as a result of their use of an enumerative strategy (a tendency to seek only information to confirm a given hypothesis), versus an eliminative strategy[2] (a tendency to search for evidence that is inconsistent with a hypothesis), and their resistance to testing a range of varied and alternative hypotheses. The results illustrated that people had a distinct aversion to trialling information that disconfirmed or discredited their original hypothesis, as most participants consistently only tested triples that conformed to their original theory (Wason, 1960). This confirmation bias was pervasive, and often, even after a hypothesis had been disconfirmed by the researcher, the participant would continue to suggest other number triples conforming to their original incorrect rule, thereby exhibiting a narrow exploration of the domain of all available triples (Wason, 1960; Klahr & Dunbar, 1988). Higher rates of success on the RDT were illustrated in individuals that exhibited a much higher rate of systematic eliminative thinking (a preference for information that disconfirmed the given hypothesis), considered a wide range of alternative hypotheses and explored a larger number of triples before announcing their initial rule. The average number of triples generated by participants who were successful on their first rule announcement was 8.0, compared to 3.68 for those individuals whose first rule announcement was incorrect (Wason, 1960). Non-solvers were also significantly more likely to follow a hypothesis-confirming strategy (termed an enumerative strategy by Wason), primarily only positing triples that invoked positive responses (Klayman & Ha, 1987, 1989; Tukey, 1986; Wason, 1960; Wharton, Cheng, & Wickens, 1993). Many other researchers have also demonstrated the importance of considering alternative hypotheses in successful hypothesis evaluation (Gale & Ball, 2009; Klayman & Ha, 1989; Mynatt, Doherty, & Tweney, 1977; Vallée-Tourangeau, Austin, & Rankin, 1995; Wason, 1960; Wickens & Flach, 1988).

Wason continued to research the phenomenon of confirmation bias and in 1968, he conducted two experiments (using his famous selection task) to explore the notion that this cognitive bias could be altered or overridden in some way. The first experiment assessed the effects of instructing participants to select, which of four cards, would render a given hypothesis false, if the values on both sides of the cards were known (Wason,

---

[2] Wason's use of the terms enumerative and eliminative are problematic. Wason measured participants' success on the task by examining the occurrences of positive and negative feedback using an enumerative/eliminative index. Without actually asking participants what they expected the result of their hypothesis test to be, it appears that his index might actually imply that an eliminative strategy was being used when in fact no negative tests were utilised.

1968). In Experiment 2, participants were given the opportunity to assess four possible alternative card combinations in relation to a given hypothesis and were informed that one of these card combinations could immediately falsify the rule. It was posited that this card combination would then be utilised more often than the other three in further tasks evaluating the truthfulness of the specified hypothesis, given that this particular contingency could falsify the condition immediately (Wason, 1968).

There was little difference in the rates of success on the task between the experimental and control groups and the treatments administered in an attempt to eliminate confirmation bias had no effect on participants' subsequent choices on either of the two experiments. Wason (1968) concluded that the poor results of the two attempted interventions further supported the theory that individuals have a deep-seated aversion to selecting information that disconfirms their ideas and beliefs as a result of their existing schemas (Wason, 1968). Wason's selection task has been the target of criticism and debate with many researchers arguing that success on the rule discovery task does not have anything to do with correctly understanding how to utilise disconfirming information and correctly making use of falsifying evidence.

1.4 Limitations of the Rule Discovery Task

Over the years, Wason's (1960) original RDT and studies utilising the RDT method have come under fire from a number of researchers (Evans, 2006; Farris & Revlin, 1989a, 1989b; Klayman & Ha, 1987, 1989; Wetherick, 1962). Previous research and investigations into individuals' methods of hypothesis testing have focused on the examination of participants' responses in terms of whether or not they are exhibiting a confirmation bias through their attempt to confirm rather than disconfirm their hypotheses. Wetherick (1962) was quick to criticise this research method, arguing that by examining only the occurrences of positive and negative feedback, without asking participants whether they expected the results of their hypothesis test to be positive or negative, meant it was impossible to correctly identify whether participants were knowingly using either an enumerative or an eliminative strategy on Wason's (1960) RDT. Wetherick (1962) proceeded to illustrate that it was in fact possible to eliminate a hypothesis by receiving a negative response to a positive test (a test that is consistent with the participant's hypothesis) or by receiving a positive response to a negative test (a test that is inconsistent with the participant’s hypothesis), thereby demonstrating that without knowledge of a participant’s expectations it is impossible to correctly identify their hypothesis-testing strategy.

A further concern of the original RDT is that it does not represent a real-life scenario or task that would require an individual to engage in hypothesis testing (Fischhoff & Beyth-Marom, 1983; Vallée-Tourangeau & Penney, 2005; Wetherick, 1962). Consequently, there are concerns surrounding the generalisation of any results obtained on the RDT to everyday situations and tasks (McKenzie, 2006). The infinite number of potential hypotheses available to trial on Wason’s (1960) original RDT would be impossible to undertake in most applied problem-solving contexts. Wason’s (1960) RDT also exhibits a situation in which the majority of initial hypotheses (‘numbers increasing by two’ being the most popular) will also fit the actual hypothesis (ascending numbers) as the set of all tests associated with the hypothesised rule are ‘embedded’ in the set of all tests pertaining to the true rule (Klayman & Ha, 1987, 1989; Wharton, Cheng, & Wickens, 1993). As Van der Henst, Rossi, and Schroyens (2002) illustrated, an individual’s hypothesis-testing ability is often quite effective when undertaken in a real-life situation that they find relevant. Problems occur when the inherent design of the research is unusual and counterintuitive to the participant. Wetherick (1962) posited that the very design of Wason’s original RDT fosters the use of an enumerative strategy in participants. Oaksford and Chater (1994) reviewed previous research and summarised the limitations of Wason’s RDT into four main points.

1. There are an infinite number of possible hypotheses available in Wason’s original 1960 experiment. In the field of science, this represents a highly unlikely scenario, as often researchers have a very limited number of plausible hypotheses available to them.
2. The triple ‘2-4-6’ provides little constraint on the available hypothesis testing space, as not much is known about the true rule behind the proposed hypothesis. In fields where a plethora of research has been undertaken, this is not usually the case.
3. Participants in the experiment are told and are aware that there is a law or rule to be discovered. This is unrepresentative of exploratory research in the field of science where researchers are not privy to this information.

4. There is an implicit suggestion in the RDT that only the true rule or theory will have any important utility. In the field of science, this is not normally the case. Many approximations of theories and ideas exist which play a vital role in peoples' understanding of a phenomenon.

Despite the illustrated limitations of Wason's original RDT, research studies following similar methods flourished after the original publication of the research in the 1960's.

1.5 Further Research Exploring Confirmation Bias in the Hypothesis Testing Tradition

Following on from Wason's experiments, a plethora of further research into the construct of confirmation bias in the hypothesis-testing tradition has been conducted (Nickerson, 1998). Much of the research has involved slight variations on Wason's (1960) original RDT, in order to explore whether or not the phenomenon of confirmation bias could be replicated under different conditions and in a varied range of situations. For example, Klayman and Ha (1989) utilised both numbers and cities in their triple construction and Tweney et al. (1980) utilised triples constructed of animal names in their respective replications of Wason's RDT. Further experiments have also served to investigate the incidence of confirmation bias in a variety of hypothesis-testing tasks (Doherty, Mynatt, Tweney, & Schiavo, 1979; Skov & Sherman, 1986; Tschirgi, 1980; Van Swol, 2007). In research conducted by Doherty et al. (1979), participants were told that they had come across a valuable archaeological find in the form of a pot and that it was important that they determine whether it had originated from Coral Island or Shell Island. The participants were then given a list of features describing the pot and pottery characteristics of the two islands. Given a short description of the pot they had found, the participants were instructed to select the data they felt was most important in deciding from which of two islands the archaeological find had originated from. The data was originally manipulated so that it appeared that the archaeological find was more likely to have originated from one of the two islands. Participants were given the opportunity to seek further information and despite disconfirming information being readily available, approximately 85% of participants primarily selected information to confirm their original hypothesis as to which island the pot had originated from (Doherty et al., 1979). In line with Wason (1960), Doherty et al. (1979) illustrated that participants' ability to choose diagnostically valuable information was poor, despite information that could have disconfirmed their hypothesis being easily accessible.

Early research exploring confirmation bias in the hypothesis-testing tradition has illustrated that confirmation bias and specifically the tendency for individuals to seek out evidence to confirm their current hypotheses or theories, exists as a robust and pervasive phenomenon in a range of situations and contexts (Wason, 1960, 1968; Doherty et al., 1979; Skov & Sherman, 1986). Tschirgi (1980) was one of the first researchers to examine hypothesis testing in both children and adults under naturally occurring conditions utilising common everyday situations. The participants were given multiple scenarios with either a good or a bad outcome and were asked to select a hypothesis test to determine which of the present variables in the scenario caused the outcome. For example, participants were given the scenario of baking a cake and informed that three alternative ingredients had been used as the normal recipe ingredients were not available. In the scenario, the cake turned out great and the baker hypothesised that it was because he used honey instead of sugar and the other two alternative ingredients were unimportant to the result. Participants were instructed to choose from three possible tests of this hypothesis, the hypothesis test that they believed would be the most useful in assessing this claim. The results from Tschirgi's (1980) research demonstrated that the older children and the adults primarily attempted to gather only evidence supporting their constructed hypothesis and were unable to focus on more than one hypothesis at a time. By contrast, the younger child participants seemed to alter items in each scenario haphazardly and did not seem to grasp the concept of only altering one variable at a time to assess the outcome. It is suggested that this variation in strategy occurred as the older children and adults had more experience with solving problems in real-life situations that had created schemas and knowledge pathways relating to how a person should react in a given situation (Tschirgi, 1980).

In line with Tschirgi's (1980) findings, Mynatt, Doherty, and Dragan (1993) posit that as a result of humans' limited cognitive capacity and working memory, people are only able to assess a single hypothesis at a time. This limited capacity results in "insufficient attentional resources to update two hypotheses at once" (Mynatt et al., 1993, p. 775). As well as only being able to assess one hypothesis at a time, Evans (2006) suggests that people are strongly attached to their hypotheses until an extremely good reason is identified to dispute it. Interestingly, the results of Tschirgi's (1980) earlier research illustrated that when participants perceived that the proposed hypothesis would result in a negative outcome, they altered their pattern of decision-making and systematically searched for the one variable that could be changed in order to disconfirm the negative hypothesis. Tschirgi (1980) suggested that this result was evidence that people are not particularly concerned about the formal structure or strategy they utilise in their manipulation of variables when testing hypotheses, people are more interested in the outcomes of manipulating a certain variable and will alter their strategy to ensure a positive result.

More recently, Van Swol's (2007) research demonstrated that confirmation bias is pervasive in real-life situations involving information processing and decision-making in groups. Van Swol (2007) illustrated that groups whose members shared a strong consensus on a topic of discussion rated information available that was congruent with their shared view as much more important than information that disagreed with the consensus. Interestingly, the effects of the confirmation bias were minimised in groups with minority members. Effectively, the minority group members forced majority group members to assess the entire hypothesis-testing domain and consider alternative hypotheses before coming to a final conclusion (Van Swol, 2007).

The results of previous research into confirmation bias in the hypothesis-testing tradition have identified that confirmation bias is comprised of several important characteristics and it encompasses more than just the tendency for a person to seek out information that confirms a current hypothesis. Confirmation bias is in fact a multi-faceted construct that consists of many aspects (Rassin, 2008; Skov & Sherman, 1986). The main identified aspects are as follows:

1. A tendency for individuals to ignore disconfirming evidence obtained in the evidence acquisition phase of hypothesis testing, in favour of data that supports a given rule (Jones & Sugden, 2001).
2. The under-weighting of evidence or data that would falsify a given rule. By contrast, when faced with confirming evidence people are quick to come to the decision that the hypothesis is correct (Fugelsang, Stein, Green, & Dunbar, 2004). Gadenne and

Oswald (1986), and Slowiaczek, Klayman, Sherman, and Skov (1992) illustrated that even with hypotheses presented to individuals that were fairly impartial in nature, data obtained in tests of these hypotheses that was consistent with the rule was considered more important than data obtained that was inconsistent with the given hypothesis or rule.

3. The biasing and interpretation of information obtained so that it satisfies the current hypothesis (Beattie & Baron, 1988; Fischhoff & Beyth-Marom, 1983; Oswald & Grosjean, 2004). Often this results in incorrect hypotheses still being considered long after they should have been discarded.
4. The tendency to ask only those questions that will favour the current hypothesis and bias this hypothesis to appear more likely than it should be. McKenzie (2006) posited that together, an individual's propensity to ask extreme questions when testing a hypothesis combined with a tendency to evaluate hypotheses with insufficient sensitivity to disconfirming information produces confirmation bias.
5. Once a hypothesis or rule has been considered likely, individuals are reluctant to change their mind in the light of falsifying evidence. Klayman (1995) and Slowiaczek et al. (1992) found that confirmation bias occurs when individuals display a persistent overconfidence in their hypothesis.

Kassin (2005) and Risinger, Saks, Thompson, and Rosenthal (2002) sum up the points as follows: regardless of what the correct hypothesis may be, once individuals form an impression or make a decision about a given situation or problem, they will search for evidence that confirms their hypothesis, interpret the evidence received in such a way as to further weight the acceptance of the rule, and even go so far as to create behavioural data that verifies their hypothesis. Any information that is gathered that is found to be ambiguous with regard to the truth of the candidate hypothesis is interpreted in a way that supports the focal hypothesis. The tendency for people to seek hypothesis-confirming information as illustrated by Wason (1960) has been replicated on a number of occasions utilising variations of the original task, all illustrating, and confirming the same results (Beattie & Baron, 1988; Nickerson, 1998). Evans (2006) suggests that the term confirmation bias has been overused and there are some situations that would be better explained by other models of hypothesis-testing and decision-making strategies. Two of the candidate competing explanations of the patterns illustrated in Wason's RDT results that have emerged in the literature, are Klayman and Ha's (1987, 1989) positive testing strategy or positivity heuristic interpretation and Farris and Revlin's (1989a, 1989b) counterfactual strategy approach.

1.6 The Positive Testing Strategy

Klayman and Ha (1987) were the first researchers to explore the positive testing strategy (PTS) in depth and propose it as an effective method of decision-making in the hypothesis-testing tradition. The researchers took into consideration Wetherick's (1962) criticisms of Wason's RDT and clarified Wason's ideas, arguing that the PTS is actually an efficient and effective method of hypothesis testing, as more often than not, it leads to a correct acceptance or rejection of an individual's proposed hypothesis (Klahr & Dunbar, 1998; Klayman & Ha, 1987, 1989; McDonald & Stenger, 1993). In following a PTS, people exhibit a tendency to test the truth of their hypothesis through the use of questions, which answered in the affirmative, would give evidence to support their current hypothesis. Klayman and Ha (1987) questioned Wason's (1960) claim that individuals make no attempt to critically evaluate their hypotheses and only ever seek out or search for evidence to confirm their theories. Because Wason's RDT did not require participants to outline to the experimenter, which hypothesis they were testing and the answer they expected to receive, it was impossible to identify whether people were using an enumerative or an eliminative strategy. Wason (1960) simply assumed that if the answer to a proposed number triple was yes, then participants must have been following an enumerative strategy, by testing, only those triples that they believed were consistent with their current hypothesis.

Klayman and Ha (1987) suggest that individuals do not only seek out evidence to confirm their hypotheses, as participants, when testing their current hypothesis, could receive negative feedback, giving them evidence that will allow them to correctly falsify their current hypothesis. The PTS posits that individuals can correctly accept or refute a given hypothesis by exploring situations in which certain outcomes are expected to occur. Tests can then be performed in these situations to positively identify if the expected results do in fact arise, or through the investigation of situations or occurrences where the expected situation has already occurred, and matching the variables in the given hypothesis to see if they prevail in the current situation (Oswald & Grosjean, 2004). Klayman and Ha's (1989) research investigated the PTS in detail utilising a variation of Wason's RDT. Identical to Wason's research, participants were given an example of a rule that the experimenter had in mind and were required to ascertain exactly what this rule was by proposing new triples and receiving feedback on whether these triples fit the experimenter's rule or not. Klayman and Ha's research then differed from earlier research, in that participants were required to announce their current best guess of the experimenter's rule at the end of every trial and there were a set number of eighteen trials to be completed. Further differences included the triples utilised, which were either numbers or cities and participants were assigned to one of three different problem conditions. Problems were classified as either being embedded (the set of all tests associated with the hypothesised rule (H) were embedded within the set of all tests associated with the true rule(T)), overlapping (the set of all tests associated with the hypothesised rule overlapped the set of all tests associated with the true rule) or surrounding (the set of all tests associated with the hypothesised rule surrounded the set of all tests associated with the true rule).

Overall, the researchers found that participants primarily engaged in a positive testing strategy, with approximately 66% of all hypothesis tests conducted by participants being classified as positive hypothesis tests (Klayman & Ha, 1989). Klayman and Ha (1989) also identified that participants in all three problem conditions made steady progress towards the true rule with over 50% of participants successfully identifying the correct solution by the end of the 18^{th} trial. Klayman and Ha (1989) found that the type of triple that participants received (either number or city) had no impact on their task performance, but those participants assigned to the embedded problem condition struggled with the task. By the end of the 18^{th} trial, 65% of those participants who did not reach the correct solution were from the embedded hypothesis condition. Klayman and Ha (1989) suggest that participants in this condition found it difficult to identify that they needed to broaden their search of the hypothesis-testing domain.

Klayman and Ha (1987) also claim that a PTS has utility in realistic hypothesis-testing situations when the probability of incorrectly accepting a hypothesis using a positive testing strategy is minimal. This claim is further supported by the research findings of Klahr and Dunbar (1988). In Klahr and Dunbar's (1988) research, participants were taught how to use a computer-controlled robot tank called 'BigTrak' and were then instructed to find out how a previously un-encountered function worked on the device. The PTS strategy was particularly effective in this task, as disconfirming evidence was more likely to be received as a result of a proposed hypothesis than confirming evidence. Regardless of the fact that participants routinely sought out information to support their hypotheses, they received disconfirming information approximately 60% of the time and as a result, participants were able to reject their incorrect hypotheses.

The PTS also has limitations (Klayman & Ha, 1987, 1989). The benefits of the PTS are not experienced in all fields and the applied literature on how to mitigate the effects of confirmation bias is small when individuals are primarily utilising a PTS. Despite the PTS being useful in situations where the set of all cases of the hypothesised rule are either, disjointed, surrounding, or overlapping, the set of all cases of the target rule, in cases where the hypothesised rule is embedded in the true rule, this can produce misleading feedback (Klayman & Ha, 1987, 1989; Rossi, Caverni, & Girotto, 2001). Following a purely PTS in such a situation will not enable participants to falsify the rule that they have hypothesised. Although the use of positive hypothesis tests can be very informative in many situations, they can only reveal false positives, but never false negatives. Specifically, positive hypothesis tests will allow a participant to identify occurrences that should be disregarded but will not allow additional occurrences to be identified that should be included (Klayman & Ha, 1987, 1989).

Similar to Wason's enumerative strategy, primary use of a PTS can lead to the overweighting of positive data and the underweighting of negative data. Paradoxically, the ability of a PTS to restrict the hypothesis testing space, which is one of the main advantages of a PTS as posited by Oswald and Grosjean (2004), is also one of the major limitations of the strategy, when the restriction of the hypothesis testing space, also closes out the true rule. Therefore, in many real-life situations where initial hypotheses are consistent with a lot of the information available in the given situation, negative feedback to a question is unlikely, and a PTS cannot be usefully employed, as it will lead to the occurrence of confirmation bias (Klahr & Dunbar, 1988; Oswald & Grosjean, 2004). Given the limitations of the PTS, Farris and Revlin (1989a, 1989b) posit that rather than using a positive testing strategy or a confirmation or disconfirmation strategy, participants engaged in hypothesis testing and evaluation in fact utilise a counterfactual strategy.

1.7 The Counterfactual Strategy

Under the counterfactual strategy, the hypothesis given by the participant in the role of experimenter is assumed false and a complement hypothesis is created that is directly opposite to the given triple. For example, if the given triple is 2-4-6, as in Wason's original RDT, a participant is likely to posit that the rule is 'even numbers', therefore a complement hypothesis of 'odd numbers' is created, being the direct opposite of the hypothesised rule and the number triple 3-5-7 is announced to the experimenter (Farris & Revlin, 1989a, 1989b). This triple conforms to the experimenter's rule of 'any three ascending numbers' and as such, it receives positive feedback. Therefore, although it may seem that an individual is attempting to confirm their own complement hypothesis by engaging in an enumerative strategy, they are in fact using an eliminative strategy i.e., by trying to find evidence for R_H by looking for evidence that would support the R_H complement. It is proposed that when used together the complement hypothesis and its confirming triple are sufficient to correctly disconfirm the original hypothesis (Farris & Revlin, 1989a, 1989b; Oaksford & Chater, 1994). With Wason's original research, it is unclear which strategy participants are employing, as the experimenter does not require the participants to inform the researcher before each test, whether they are expecting a positive or negative feedback response to their proposed triple (Wetherick, 1962).

Farris and Revlin's (1989a) research followed Wason's (1960) original RDT method. The results were similar to Wason's in that only 37% of the participants announced the correct rule on their first trial. Interestingly, the researchers illustrated that 70% of participants who were able to correctly state the rule on their first attempt, did so without using any negative tests. Farris and Revlin (1989a) posit that the use of a disconfirming strategy, although sufficient to correctly identify the true rule, is not necessary. The counterfactual strategy explanation was effective at distinguishing those participants who solved the RDT from non-solvers; in contrast, the use of Wason's eliminative strategy explanation was not useful in distinguishing solvers from non-solvers (Farris & Revlin, 1989a). Farris and Revlin (1989a, 1989b) conclude that rather than believing that people always follow an ineffective method of decision-making in hypothesis testing, individuals often engage in relatively sensible, rational, and effective reasoning processes. In their later research, Farris and Revlin (1989b) explored the strategies utilised by participants on the RDT, to distinguish differences between those who were successful on the task and those participants who were not. Results indicated that participants who solved the RDT on their first rule announcement engaged in the use of a counterfactual strategy twice as often as those who were unsuccessful on their first rule announcement. Solvers also generated approximately 53% more trials of their proposed hypothesis before announcing a rule (Farris & Revlin, 1989a).

Farris and Revlin's (1989a, 1989b) counterfactual strategy has been criticised by Oaksford and Chater (1994) for a crucial oversight which seriously limits its applicability. They note that the positive response received by the participant from the experimenter when testing a complement hypothesis suggests that the original hypothesis of 'even numbers' is false and the complement of 'odd numbers' is plausible. However, logically this does not make sense, as the rule that the participants are attempting to unearth is meant to share similarities with the original triple. The complement hypothesis of 'odd numbers' cannot plausibly be the correct rule, as none of the number's in the seed triple were odd numbers (2-4-6). Hence, the odd numbers rule must also be false.

Thus, Oaksford and Chater (1994) propose a modified version of the counterfactual strategy, that they call the iterative counterfactual strategy that overcomes this problem. In contrast to Klayman and Ha's (1987) PTS, the iterative counterfactual strategy works extremely well in situations where the hypothesised rule is embedded in the true rule. They suggest that because both the seed triple of 2-4-6 and the complement 3-5-7 and their respective hypotheses of 'even numbers' and 'odd numbers' are both known to be instances of the true rule, the participant can engage in an enumerative strategy to propose a new set of hypotheses and complements based on a property shared by both triples. This strategy continues until a negative response is received as a result of a complement test, suggesting that the hypothesised rule is plausible. Participants then enter into the 'positive sub-loop' phase of testing where they generate further triples to receive further positive feedback before announcing their rule. The 'positive sub-loop' posited by Oaksford and Chater (1994) is analogous with Bayesian theory which posits that individuals gradually come to accept a hypothesis as true, as a result of an accumulation of evidence supporting the theory (Nelson, 2005). Similarly, research by Müller, Garcia-Retamero, Cokely, and Maldonado (2011) suggests that the same process occurs when people are faced with an accumulation of disconfirming evidence, allowing them to adjust their beliefs and select correct solutions despite strongly held causal beliefs.

Both Oaksford and Chater (1994) and Tukey (1986) observe that individuals do not often employ only one strategy of hypothesis testing in any given situation. Previous research has also suggested that it is questionable whether the results of prior RDT research are generalisable to all, or even many real-world scenarios (Raab & Johnson, 2007). The situational factors of a scenario and an individual's previous experience with a present problem, all serve to influence the strategies that people employ and specific problem solving situations provide different constraints on reasoning and thus different strategies may be more, or less effective. For example, in a world where H is never embedded in T, the PTS will work fine almost every time and in Wason's rule discovery task there are barely any constraints, which is very dissimilar to hypothesis-testing in science and real-life scenarios. It is therefore important to investigate the incidence of confirmation bias in research studies that have employed realistic investigations of hypothesis testing and evaluation. Of the studies that have been conducted, the fundamental findings of earlier RDT research have been replicated.

Chapter Two

Realistic Applications of Hypothesis Testing and the Incidence of Confirmation Bias

2.1 Realistic Applications of Hypothesis Testing in the RDT

Mynatt et al. (1977) pioneered research into the exploration of confirmation bias using scenarios that more accurately represented real world situations. The researchers provided the field with more lifelike applications of hypothesis testing and rule discovery in simulated environments, which represented real problem-solving events requiring participants to undertake tasks involving realistic hypothesis evaluation and testing. In their first experimental study, Mynatt et al. (1977) instructed participants to posit hypotheses about the rules that existed in a complex, artificial environment. Participants gained evidence of the scenario's features by firing particles directed at any part of a computer screen and observing how the particles interacted with the objects present. By following the behaviour of the fired particles, participants postulated hypotheses about the environment that would account for their motion. The task involved a complex interaction between many programme characteristics, and was constructed in such a way, that participants' initial trial uses of the system suggested that certain characteristics of the environment accounted for the particle's motion (Mynatt et al., 1977). After the initial task trial, participants constructed possible hypotheses to evaluate and all of their subsequent actions during the remaining experimental tasks were monitored for the display of strategies to either confirm, or disconfirm their original hypotheses (Mynatt et al., 1977).

One of these tasks was the paired screen choice task. In this task, participants were given the opportunity to choose from two screens, the screen that they felt would be the most useful to them in further testing the truth of their initial hypothesis. One screen displayed features congruent with their hypothesis and the second screen displayed novel objects and conditions. A positive test was indicated by the participant choosing the screen congruent with their original hypothesis, as it indicated a tendency to search for evidence to confirm, rather than disconfirm their hypothesis (Mynatt et al., 1977). Overwhelmingly, participants exhibited a hypothesis-confirming response pattern with 71% of participants selecting the positive test. Mynatt et al's. (1977) results illustrated that the majority of participants had an aversion to trialling alternative hypotheses and the researchers posited that their results indicated that confirmation bias is not just applicable in abstract tasks, it can also occur in simulated scientific environments.

Mynatt, Doherty, and Tweney (1978) followed on from their earlier research in 1977, with the re-design of their original particle motion environment. The new system simulated a very complex and dynamic environment containing 27 different fixed objects, some of which influenced the path and direction of a moving particle. Participants were instructed to come up with hypotheses about the environment and run tests of these hypotheses in order to discover the rules governing the movement of the particle in the system. Despite all of the participants being intelligent university students majoring in various science fields, none of the sixteen participants were able to find the correct solution (Mynatt et al., 1978). As a direct consequence of the complex nature of the task, participants received many disconfirming responses to their trialled hypotheses throughout their exploration of the artificial environment. Despite being faced with a plethora of instances of disconfirming evidence, approximately 70% of participants struggled to abandon their hypotheses, with some participants discarding incorrect hypotheses momentarily, only to come back to them later. In Mynatt et al's. (1977) earlier research, participants were quick to abandon hypotheses in light of disconfirming evidence when this evidence was made salient to them, with approximately 91% of participants who had selected the most obvious, yet incorrect solution, altering their initial incorrect hypothesis to a correct or partially correct hypothesis, when faced with unambiguous disconfirming information. Mynatt et al. (1978) argue that the difference in complexity between the research tasks of 1977 and 1978 may explain the difference in their participants' ability to abandon incorrect hypotheses in the face of disconfirming information.

Klahr and Dunbar's (1988) research discussed earlier, which required participants to create and test hypotheses about the role of an un-encountered function on the 'BigTrak' robot, was also one of the first pieces of research to systematically explore the use of different hypothesis testing strategies in simulated real-world environments. As well as supporting the aspects of Klayman and Ha's (1987, 1989) PTS research, Klahr and Dunbar's (1988) findings also replicated many of the findings of Mynatt et al. (1978). Despite participants being able to correctly disconfirm incorrect hypotheses approximately 60% of the time using positive tests, it was discovered that when participants were confronted with evidence that disconfirmed their hypotheses, they did not immediately discard them. In Klahr and Dunbar's (1988) first study, participants discarded hypotheses immediately following disconfirming evidence in less than 50% of all cases. The researchers speculate that the task complexity, participants' inability to think of alternative hypotheses, and a belief that not enough disconfirming evidence had been accumulated against the current hypothesis, all contributed to this outcome.

Teasley, Leventhal, Mynatt, and Rohlman (1994) also investigated Klayman and Ha's (1987, 1989) positive testing strategy to see if the findings could be replicated in a software testing scenario. Participants were tasked with testing a computer programme whose hypothesised function was to read in a line of text and output the number of times each lower-case letter occurred. Participants were also informed that the programme accepted a maximum of 250 characters. The task was to test the programme to check it was in fact running to these specifications. Participants were scored according to how many +S tests (positive tests of the programme specifications) and how many –S tests (tests testing outside the proposed programme specifications) were utilised throughout the experiment. In accordance with Klayman and Ha's (1987) ideas, strong evidence of the positive testing heuristic was obtained, with participants in Experiment 1 performing three times more positive tests than negative tests, which indicated a greater coverage (exploration) of the domain of all possible positive tests.

In their second experiment, Teasley et al. (1994) examined the effect of making the domain of possible negative hypothesis tests or –S tests more significant by varying the detail specifications of the software-testing task. Three different possible types of specifications were given to the participants, minimal specifications, +S only specifications and specifications containing information on both +S and –S tests. Similar to their first experiment, there was a strong indication of the use of a positive testing strategy among participants and again participants' coverage of the negative testing domain space was lower than the coverage of all possible positive tests. Overall, both the number of +S and –S tests trialled increased with increasing levels of participant experience (Teasley et al., 1994). The maximum programme specification group who received information on both –S and +S tests, utilised a larger number of –S tests. However, even the advanced testers in this group did not achieve full coverage of all available testing domains, covering approximately 90% of the positive testing domain space and only attaining 60% coverage of the negative testing domain. There were no significant differences in task performance between the minimal specification group and the +S specification group in the amount of +S and –S tests evaluated. Teasley et al. (1994) suggested that, as individuals inherently use a positive testing strategy, explicitly stating the use of +S tests, does not achieve a change in an individual's inherent strategy.

Vallée-Tourangeau and Penney (2005) carried out research to explore the effects of externally representing the triple features inherent in Wason's (1960) original RDT, on a participant's task success. The researchers felt that prior research into the RDT had not been representative of real-world problem solving situations, in that participants had only had access to internal representations of the task stimuli whereas during real-world problem solving, people have to examine and manipulate external cues to test their hypotheses (Vallée-Tourangeau & Penney, 2005). Participants were instructed to discover a rule that the experimenter had in mind, similar to Wason (1960), but instead of the rule being verbalised, the participants were presented with three dice, visually illustrating the number triple, 2-4-6. Participants were able to physically manipulate the dice to suggest further number triples for the experimenter to provide feedback on. It was hypothesised that an external representation of the rule would allow participants to be more creative in their exploration of the hypothesis space and would also engage the subjects with the task for longer, allowing them to generate more triples and subsequently result in higher success rates for the task overall (Vallée-Tourangeau & Penney, 2005). Of the participants in the dice condition, 66% announced the correct rule, in contrast with only 19% of participants in the control condition who completed the original version of Wason's (1960) RDT task. The results illustrated that an external representation of task stimuli greatly aided performance on the RDT. Task solvers also tested significantly more triples, representing a broader exploration of the domain of possible triples and they also tested significantly more negative or disconfirming triples.

Vallée-Tourangeau and Penney's (2005) second experiment replicated the results of Experiment 1, utilising external stimuli that participants would not have encountered commonly in everyday life. The results illustrated that a participant's familiarity with an object was less important to their level of task success, than the actual external representation of the task itself, as levels of task success remained high when the familiar dice were removed and replaced with unfamiliar hexagonal chips. Vallée-Tourangeau and Payton (2008) further replicated the earlier findings and argued that the accumulation of analogous results from numerous studies conducted on external representation of task stimuli suggest that if an experimental scenario is closely mirrored to a real-life task or problem-solving situation, where features can be interacted with and are externally represented, then success rates on the RDT will increase and participants will be much better equipped and able to discover the true rule or solution to the problem. The outcomes of the above research studies indicate that in real-world situations, people may not be as terrible at carrying out efficient and effective hypothesis tests as the original rule discovery task experiments suggest.

2.2 Susceptibility to Confirmation Bias

Dardenne and Leyens (1995) argue that the use of a hypothesis-confirming strategy is in fact an adaptive and social skill that humans have developed in order to aid them in quick and efficient hypothesis evaluation and testing. In recognition of the robust and pervasive nature of confirmation bias, researchers have begun looking into the effects of different interventions and hypothesis testing strategies on participants' success on the RDT, in an attempt to identify ways to reduce the occurrence of confirmation bias. Research in this field has met with some success, yet it appears that the effects of explicit instruction about confirmation bias and the strategies that can be utilised to avoid it are minimal at best (Kareev, Halberstadt, & Shafir, 1993). A study by Lord, Ross, and Lepper in 1979 explored the effect of providing disconfirming and confirming evidence to participants who were either opposed to or were advocates of, capital punishment. After participants received evidence either for or against the practice, they were asked to rate any change in their attitude towards capital punishment. Lord et al. (1979) found that regardless of whether the evidence received was for or against the participant's opinion, the evidence served to polarise participants' positions on the topic and participants' views on capital punishment were more extreme at the conclusion of the research, than when the research began. Lord et al. (1979) postulated that far from being effective in reducing confirmation bias, the occurrence of disconfirming information was actually counterproductive.

In agreement with Lord et al. (1979), Poletiek (1996) empirically explored her claim that asking participants to follow a disconfirming strategy is paradoxical, as asking participants to develop a best guess hypothesis and then attempt to disconfirm it, assumes that a better hypothesis exists and participant's original best guess hypothesis *cannot* be their best guess. The researcher instructed participants to perform falsifying tests of a given hypothesis in an RDT using Wetherick's (1962) method, which involved asking participants to explicitly state their hypothesis for each test. The participants were further instructed to perform falsifying tests and to give their best guess hypothesis when announcing the rule. Both their success on the task and their ability to correctly interpret and follow the 'falsify' instruction were assessed. Poletiek (1996) found that there was in fact a negative relationship between a participant's ability to discover the rule in the RDT and the incidence of using a falsifying strategy and regardless of the testing strategy employed by the participants (positive or negative), they still expected the evidence they received to confirm their original best guess hypothesis. Poletiek (1996) claimed that people struggle to follow instructions to try and disconfirm their hypotheses and as a result, individuals persist with strategies of hypothesis testing that are confirmatory in nature.

Similarly, Mynatt et al. (1977, 1978) argued that instructing participants to utilise a disconfirmatory strategy is not an effective method for reducing a participant's susceptibility to confirmation bias. Their research conducted in 1977 assigned participants to one of three experimental conditions which were then administered variations on the original RDT task instructions: (a) group one was given instructions to attempt to confirm their hypothesis, (b) group two was given explicit instructions to attempt to disconfirm their hypothesis, and (c) group three received general instructions to test their hypothesis without the experimenters specifying a method. Mynatt et al. (1977) found no significant differences on task success between the different instructional methods. In 1978, Mynatt et al. again attempted to identify methods that would reduce the occurrence of confirmation bias in participants. This research was conducted with two experimental groups, one of which received no special instructions on hypothesis testing and the value of falsification, and one group who received extensive information on hypothesis testing and the value of falsification. Again, no significant differences were found on task success between participants from the two groups. Mynatt et al's. (1977, 1978) research illustrated that the administration of explicit instructions to participants to follow a disconfirming strategy did not increase the likelihood that participants would follow a hypothesis-disconfirming strategy, nor did they improve participants' subsequent evaluation of any disconfirming information received during the course of their hypothesis testing process.

Research by Tweney et al. (1980) discovered that the greatest increase in peoples' ability to solve a problem-solving task arose when participants were informed that there were two rules to be discovered. This scenario was termed a dual goal (DG) condition. A scenario containing a single goal (SG) condition was identical to the original rule discovery task where there exists one rule for the participant to try and discover and they are informed simply by the researcher whether or not their generated triple conforms to the rule that the researcher has in mind. In a DG condition, there are two rules that are to be discovered and triples proposed are assigned to one of two categories. In Tweney et al's. (1980) research, these two categories were labelled DAX and MED. After proposing possible triples, participants then received feedback on whether a proposed triple was a DAX triple or a MED triple. MED triples were complement triples of DAX, so that any triples that did not conform to the DAX category were automatically considered MED rules. Feedback received in the DG condition, which in the SG condition would have been discarded as incorrect, was now considered important as it could be related to the alternative rule. By contrast, to the SG condition, which prevented participants from using all of the information available to them, participants in the DG condition effectively received two different types of confirmatory information, as feedback received, either confirmed characteristics about the DAX rule, or confirmed features about the MED rule. Participants in the DG condition compared to participants in the SG condition, carried out an increased number of tests, and utilised a greater variety of different tests, thereby increasing their chances of finding the true rule. The DG effectively requires participants to broaden their hypothesis-testing space (Tweney et al., 1980). Utilising a DG manipulation of Wason's (1960) original RDT in their fourth experiment dramatically increased participants' ability to correctly identify the true rule for DAX (which was identical to Wason's (1960) 'true rule'), with over 60% of their participants solving the task correctly on their first rule announcement (Tweney et al., 1980; also see, Gale & Ball, 2006, 2009).

Tweney et al. (1980) argued that simply requesting participants to consider dual hypotheses does not work. In order for the DG condition to promote increased rates of task success, participants must be able to critically evaluate and seek evidence for both hypotheses simultaneously and understand that the two hypotheses are complementary. Wharton, Cheng, and Wickens (1993) conducted three experiments on the effects of DG conditions on task success. Experiment 1 required participants in the DG condition to label each of their hypotheses and the feedback received, after a rule announcement, as either

DAX or MED, but participants were not made aware of the complementarity of the two rules. Results illustrated that with the relationship between DAX and MED left ambiguous, the task performance of the participants in the DG condition was no better than the task performance of participants in the SG condition. To explore the importance of participants being informed of the complementary relationship between DAX and MED on subsequent task performance, participants in the DG condition in Experiment 2 were told that the DAX and MED rules were complementary and the relationship was clearly explained to the participants if they were unsure. Results from this experiment illustrated that participants in the DG condition performed significantly better than participants in the SG condition, with 43% of DG participants announcing the correct rule on their first announcement, versus only 14% of participants in the SG condition (Wharton et al., 1993). Experiment 3 required all participants to test at least five triples before announcing their first rule. The results indicated that this manipulation did not improve the task performance of participants in the SG condition to the level of participants in the DG condition, the researchers suggested that it was the type of information being received by the DG participants that was important, and not the quantity of information received (Wharton et al., 1993). Similar to Tweney et al. (1980), Wharton et al. (1993) concluded that participants in the DG condition performed significantly better on the adapted RDT when it was made explicit to them that the DAX and MED rules were complementary.

The success of participants in DG tasks has been replicated on many occasions. Further research by Vallée-Tourangeau, Austin, and Rankin (1995) illustrated that participants had a success rate on their hypothesis-testing task of 81% in the DG condition versus a success rate of only 41% in participants in the SG condition. In contrast to Wharton et al. (1993) and Tweney et al. (1980), Vallée-Tourangeau et al. (1995) did not find that it was necessary for the complementarity of the two rules to be made explicit in order for the DG task to have any significant effect on participants' task success. In line with earlier findings by Wason (1960) and Klayman and Ha (1987, 1989), the researchers claimed that success on the task was likely due to participants' production of a large number of possible triples and a larger number of triples that produced negative feedback (Vallée-Tourangeau et al., 1995). More recently, research conducted by Gale and Ball (2006, 2009) has provided evidence for the vital role of 'contrast class cues' in the success of participants on the RDT. Gale and Ball (2006, 2009) suggest that testing and evaluating triples that are in direct contrast with the most likely rule hypothesis (e.g., testing a number triple of 'descending numbers' in contrast to Wason's (1960) original RDT rule of ascending numbers), is extremely useful in assessing the boundaries of the true rule domain space. Despite the replicability of DG research, Penner and Klahr (1996) argue that it is limited in its applicability to real-world scenarios as it is highly unlikely that people will commonly experience situations where opposing hypotheses are mutually exclusive or represent an exhaustive representation of the possible solution set of the domain space.

Schwind, Buder, Cress, and Hesse (2012) utilised a web-based search engine to investigate whether the recommendation of specific search results to participants could be used to reduce the incidence of confirmation bias. The researchers created a fake list of eight websites that appeared as search results in the topic of neuro-enhancement, four of these websites supported neuro-enhancement and the remaining four websites were against the topic. Participants were given some brief information on neuro-enhancement and were asked to indicate their opinion for or against the topic, to allow the researchers to assess any latter change in opinion. In the experimental condition, one of the eight arguments for or against neuro-enhancement was highlighted with an orange box and was explicitly recommended to the participant. Participants were then asked to select the one piece of information contained in the search results that they would like to further explore. Schwind et al. (2012) hypothesised that, contrary to Lord et al's. (1979) earlier research, providing individuals with explicit preference-disconfirming recommendations significantly reduced the incidence of confirmation bias. This was clearly illustrated in their findings - in the experimental group that received an explicit preference-disconfirming recommendation, only 15% of participants experienced confirmation bias, whereas in the experimental group who received a preference-confirming recommendation, 51% of participants experienced confirmation bias. The researchers' argued that their findings suggest that although the mere existence of disconfirming information may not be sufficient to encourage a change in theory, if evidence is made explicit as a salient point to the individual, it will be taken into consideration in the gathering of information in a hypothesis-testing task (Schwind et al., 2012). This notion is further supported by earlier research conducted by Wason (1968) and Jones and Sugden (2001) who both postulated that simply providing people with evidence of disconfirming information is not effective as a means of reducing confirmation bias, as the effects of confirmation bias are robust and persistent. However, participants are almost always able to make the correct judgement in the RDT if sufficient evidence exists to disconfirm the original theory and this evidence is made salient. Unfortunately, in most real-world situations, people will not always have the advantage of having disconfirming evidence made explicit to them in the information acquisition stage of hypothesis testing. Often people will have to sift through large amounts of data and select those pieces of evidence that they deem most important to the task at hand. Without the luxury of having important information highlighted for them, it is likely that this information will not assist in the correct dismissal of an incorrect hypothesis as Lord et al. (1979) showed.

Cook and Smallman's (2008) approach to the investigation of de-biasing methods moved away from the popular method of training participants in hypothesis testing strategies and guiding individuals on how to explicitly avoid biases. The researchers investigated two types of de-biasing approaches in the field of naval information intelligence assessment that they developed in line with two of the prevalent explanations of confirmation bias: (a) confirmation bias results from internal psychological processes that underlie judgement and decision-making processes, (b) confirmation bias results from external factors inherent in the task environment. Using JIGSAW (Joint Intelligence Graphical Situation Awareness Web), the researchers investigated the effects of the two different de-biasing approaches on the reduction of confirmation bias. The first approach involved the manipulation of the way in which evidence for or against the given hypothesis was presented on the participant's computer screen and the second de-biasing approach involved the distribution of information to the participants regarding the assessment of the current evidence by the participant's peers. These two approaches imitated the internal versus external explanations of confirmation bias. During the experiment, two independent variables were manipulated to represent the two de-biasing techniques. These involved the manipulation of the evidence layout format (graphical versus text) and the manipulation of the source of the evidence assessment (analyst's own views versus peer analysts' assessments). Across all conditions, participants displayed an inherent bias to select evidence that supported their hypotheses, with only 19% of evidence selections across all conditions being disconfirming evidence. Evidence illustrated graphically was the only manipulation that reduced the incidence of confirmation bias, although even this effect was minimal, increasing the likelihood that an individual would select the more effective disconfirming evidence by only 6%, in comparison with the scenarios where evidence was presented in text only.

Further research has investigated other possible strategies that may mitigate the effects of confirmation bias other than the explicit incidence of hypothesis-disconfirming information. Dawson, Gilovich, and Regan (2002) found that participants were four times more likely to select a card combination that invalidated the task rule when the current rule posed a possible detrimental personal effect, than individuals who did not experience this self-preservation motivation. These results are similar to Tschirgi (1980) who demonstrated that both adults and children alike were motivated to search for disconfirming information when the truthfulness of the given hypothesis was deemed likely to result in an undesirable outcome for the participant. It appears that the susceptibility to confirmation bias can be overridden by peoples' strong inherent nature to protect and preserve their wellbeing (Dawson et al., 2002; Tschirgi, 1980).

Confirmation bias in its many forms is one of the most researched biases in the field of psychology (Nickerson, 1998). Mercier and Sperber (2011) noted that despite the abundance of research available, much has neglected to identify any significant individual differences, de-biasing therapies, or hypothesis-testing strategies that have been shown to significantly influence an individual's susceptibility to confirmation bias. Repeatedly, studies have shown that it appears the majority of people are prone to confirmation bias but little research has explored why. Despite the incidence of confirmation bias in research participants, it does not appear to be as a result of the utilisation of any one particular strategy (Beattie & Baron, 1988). Rassin (2008) hypothesised that there exists 'a priori' individual differences in confirmation proneness and research to date has neglected to explore the influences of these individual differences that exist before experimental conditions are put in place. Rassin (2008) further argues that individual differences may predispose individuals to confirmation bias and exploration into this field will advance knowledge into effective solutions for correcting this ingrained bias in decision-making. Hart, Albarracín, Eagly, Brechan, Lindberg, and Merrill (2009), conducted a meta-analysis of 91 empirical research studies in the field of decision-making and confirmation bias to explore the notion that certain individual differences exist that influence the susceptibility to confirmation bias. The researchers demonstrated that over the empirical research studies analysed there were several characteristics that commonly arose as being identified as having a significant impact on the susceptibility of an individual to confirmation bias. Quality of information presented, commitment to pre-existing knowledge or beliefs, relevance of the current hypothesis to an individual's value system, and closed-mindedness, were all positively correlated with the incidence of confirmation bias.

2.3 The Effects of Experience on Confirmation Bias

An individual difference that has been widely researched is the effect of experience on an individual's ability to solve problems, make decisions, and engage in hypothesis testing and evaluation. Despite the abundance of theoretical information available, the results presented are conflicting. On the one hand, some research suggests that experts can fail on tasks that can be easily solved by novices, especially in situations containing unfamiliar or novel task stimuli (Hecht & Proffitt, 1995; Lewandowsky & Kirsner, 2000; Wiley, 1998). However, further research and theories on expertise also exist demonstrating that the superior knowledge of experienced individuals allows them to quickly and efficiently evaluate alternative hypotheses and select the correct course of action. Therefore, the more experienced an individual is, the less likely they are to succumb to the effects of cognitive biases such as confirmation bias (Chassy & Gobet, 2011; Ericsson & Charness, 1994; Ericsson & Lehmann, 1996).

Research conducted by Wiley (1998) explored the effects of varying levels of baseball knowledge on peoples' susceptibility to confirmation bias. Wiley was interested in the effects of prior domain knowledge on the activation of mental sets and the promotion of fixation in a creative problem-solving task related to baseball. The research also looked at whether or not any latent effects of domain knowledge that were identified could be eliminated by instructing participants to actively attempt not to rely on any prior baseball knowledge in completing the task. Across three different experiments, the results clearly demonstrated that participants with a higher degree of baseball-related domain knowledge were far more susceptible to fixation on an area of the hypothesis testing space. Wiley postulated that this occurred as a result of the activation of mental sets that were not applicable to the novel task and as a result individuals with the most experience and knowledge were the least likely to be able to solve the problem. Wiley posits that the higher amount of domain knowledge biases individuals towards providing a specific domain-related solution. This bias then prevents a broad search of the evidential domain and subsequently reduces an individual's ability to find the appropriate solution. Wiley recognises that generally, more experience will allow an individual to have access to a much greater set of possible problem solutions and often they will be able to solve problems much more efficiently and accurately than beginners. However, when experienced individuals are faced with novel tasks that require a further search of the domain space, where existing schemas are inappropriate, experience may hinder an expert's ability to solve a problem.

Similar to Wiley (1998), earlier research conducted by Hecht and Proffitt (1995) replicating Luchin's water-level task, found that the most experienced waitresses and bartenders performed more poorly than members of the general population, when asked to draw on a picture of a tilted container, the surface orientation of the liquid filling the vessel. The researchers found that the incidence of error increased as a function of years of experience with waitresses and bartenders indicating that the level of the water would deviate from horizontal significantly more than those participants with no professional experience of liquid surface-orientation (Hecht & Proffitt, 1995). Waitresses and bartenders scored incorrectly on the task 68% of the time, whereas, the less experienced group of housewives and bus drivers scored incorrectly only 48% of the time (Hecht & Proffitt, 1995).

Lewandowsky and Kirsner (2000) also explored the effects of expertise and prior domain knowledge through their investigation of experienced individuals' ability to predict the spread of bush fires in situations containing unfamiliar conditions. The scenario involved two primary predictor variables relating to the spread of the fire that were in direct opposition e.g., a fire spreading uphill against prevailing light winds. The researchers predicted that experienced individuals would have extensive domain knowledge regarding the correct course of action to take in scenarios containing either of these primary predictors alone, but not in situations containing both primary predictor variables acting in opposition. Lewandowsky and Kirsner's (2000) findings illustrated that experts were able to accurately describe and solve problems when the scenario characteristics matched their prior knowledge schemas. When information obtained contradicted participants' prior experiences on how to react to the primary predictor variables experienced in the uncharacteristic bush fire scenario, experts struggled to solve the problem.

More recently, Bilalic, McLeod, and Gobet (2008a, 2008b, 2010) extensively researched the implications of expertise on decision-making in chess players, proposing the existence of a mental set or *Einstellung* effect. The researchers postulated that the

Einstellung effect is a fixation of thought or an inability to assess and evaluate alternative hypotheses as a result of an individual's prior experience and familiarity with a context or situation. The power of this effect is posited to be very strong and Bilalic et al. (2008a, 2008b) have repeatedly demonstrated its power in expert chess players. Bilalic et al. (2010) propose that people are often oblivious to the fact that they are being affected by it and it is not until something goes wrong, or a situation characteristic does not match a person's cognitive assumptions, that issues arise, as the effect can completely prevent the consideration of alternative hypotheses.

Bilalic et al. (2008a) compared the playing ability of experienced chess players to weaker chess players over a range of chess scenarios. Their results indicated that participants struggled to consider new alternative solutions to scenarios where a familiar pathway existed, despite the familiar solution not being optimal. The ability of the experienced chess players suffered significantly; with the occurrence of the *Einstellung* effect reducing the performance of the expert chess players to a level of players approximately three standard deviations lower in skill level. Further research by Bilalic et al. (2008b) using eye-tracking equipment found that, despite individual participants in the experiment stating that they were trying to look for a shorter solution when given a two solution problem where one solution was well known, the eye movements as recorded by the eye tracking equipment clearly illustrated that despite this assertion, the individual participants continued to focus on the chess board squares and chess pieces associated with the common known solution (Bilalic et al., 2008b).

Later research by Bilalic et al. in 2010 further supported their earlier findings, which suggested that the expertise level of a chess player did not reduce their susceptibility to cognitive biases such as confirmation bias and the *Einstellung* effect. The researchers found that once the familiar solution to the chess problem was entirely removed from the chess scenario, all of the expert participants were able to find the alternative, optimal solution. Bilalic et al. (2010) also demonstrated that the effect of being shown the original familiar problem before it was removed from the situation, served to impact on the time taken for the experts to solve the task. The control group of experts were not administered the familiar solution to the problem first and were of a same skill level to the experimental group of experts. Despite the similarities in skill level, the control group found the ideal solution in 37 seconds, which was approximately half the time it took the group of experts who had been exposed to the two-solution problem (Bilalic et al., 2010). Bilalic et al. (2008a, 2008b, 2010) suggest that the findings from their research studies demonstrate how the activation of a mental set or schema can prevent or restrict the evaluation of alternative solutions and how evidence contradicting any activated schemas will not be accepted, as it does not fit with the original hypothesis.

In contrast, additional research and existing theories in the field of expertise have illustrated that extensive learning and specialisation in a field actually allows an individual to more easily adapt and broaden their skill set for dealing with situations and problems (Adams, 1993; Bilalic et al., 2008a; O'Hare et al., 2008; Raab & Johnson, 2007). Being an expert in a specific field requires a lot of dedicated practice, learning, and training. Ericsson and Charness (1994) argue that people will only reach the level of an expert when they have completed at least a decade of high levels of daily practice. Individuals with a large amount of experience, who are deemed experts, are able to call upon their wealth of knowledge that has accrued over time, to respond accurately and quickly in situations that are dynamic and require quick decision-making (Ericsson & Charness, 1994; Ericsson & Lehmann, 1996). Experienced individuals differ in this respect from novices, as the knowledge schemas relating to how they should react in a familiar situation are already in place, allowing for a much more efficient search of the solution space (Chassy & Gobet, 2011; Wiley, 1998).

Research conducted by Bilalic et al. (2008a) suggested that high levels of experience adversely affected individuals' performance on a chess scenario when the occurrence of a familiar game pathway resulted in experienced players not being able to recognise alternative solutions. Interestingly, the researchers also noted that the *Einstellung* effect had absolutely no effect on the strongest chess players, with the Grand Masters (skill level 5.5 SD's above the mean) having no trouble finding the optimal solution to a two-solution problem compared to the other expert players (skill level 3-5 SD's above the mean), who did struggle with this task. It is suggested that during the hypothesis-testing task, schemas relating to the familiar situational features were activated in experienced individuals, then in reviewing their original hypothesis, those individuals with extremely high levels of expertise were also able to identify the alternative, more optimal solution (Bilalic et al., 2008a). This was clearly illustrated by the significant score decrease of the expert participants by an average of 100 points and their subsequent reduced ability to solve the task, in contrast, the Grand Masters participating in the experiment solved all available solutions with no subsequent drop in skill level.

In a realistic application of hypothesis testing in an RDT, research conducted by Teasley et al. (1994) as discussed earlier, investigated participants' susceptibility to confirmation bias on a task involving the evaluation and testing of a piece of software. In exploring the results by expertise level, little difference was found between experts and novices across the number of +S tests (positive tests of the programme specifications) utilised, but experienced participants utilised a significantly larger number of –S tests (tests testing outside the proposed programme specifications), M = 2.70 versus M = 1.12. All participants were found to test a much greater range of positive tests than negative tests and only those individuals with greater levels of experience illustrated more coverage of the possible alternative negative hypothesis tests indicating that expert participants were less likely to rely on the use of positive tests to assess the truth of their hypotheses (Teasley et al., 1994).

Research by Raab and Johnson (2007) explored the performance of experts, near experts and non-experts in the sporting field of handball utilising eye-tracking equipment to verify the extent to which participants were engaging in a full and effective information search of the possible solution space. Participants were tasked with generating alternative solutions to fifteen handball scenarios that were presented by video, which illustrated the setup of an attack play. Once an initial solution and alternatives had been presented, participants reviewed their alternative moves and selected their final solution. Raab and Johnson (2007) showed that level of expertise had a significant effect on the generation of alternative solutions and the final solution chosen to the attack task. Expert participants generated higher quality alternative hypotheses and their final solution choice was of a significantly higher quality than the near expert and non-expert groups. It was also apparent that the expert participants were not as prone to spatial fixation when searching for visual information more often than both the near expert and non-expert groups.

2.4 Expertise in Aviation

In the field of aviation, research has demonstrated that more experienced pilots are able to draw on their available knowledge more quickly than novices, who often have to spend more time searching and evaluating all of the factors in a situation to determine those that are important and relevant before they can make a decision. Adams (1993) claimed that the more experienced a pilot is, the higher their capacity for enhanced decision-making will be.

> Information is stored in terms of schema based upon experience and training. The pilot uses pattern recognition and dynamic interrelationships among objects, situations and results to associate, integrate and interpret related knowledge instead of the static, linear thinking of the novice (Adams, 1993, p. 9).

The organisation of situational data requires pilots to filter, prioritise, and structure the available information quickly and accurately. Experienced pilots rely on their prior knowledge and experience to identify the most salient pieces of this information that will allow them to understand and make sense of the current issue or situation (Adams, 1993). The pattern of evidence, the existing situational characteristics, the action taken and the outcome are then grouped together to strengthen existing schemas or form new ones, that are then encoded and stored in the pilot's long-term memory (Wickens & Flach, 1988). Experienced pilots can then refer back to these chunks of knowledge later when the same group of features are experienced again. These knowledge chunks or templates of situational characteristics lead to pattern recognition, allowing the quick and efficient activation of appropriate responses to scenarios (Chassy & Gobet, 2011; Simon, 1974).

Madhavan and Lacson (2006) reviewed existing experimental data in the field of aviation decision-making in inclement weather conditions in an attempt to identify the external and internal factors affecting the efficacy of pilot decision-making processes in the cockpit. Their review concluded that less experienced pilots were more likely to over-estimate their abilities in the assessment of situational factors and variables in comparison with more experienced pilots. More experienced pilots were found to be less prone to decision-making biases such as confirmation bias, the overconfidence bias, and the salience bias (which occurs when people are misled by salient but incorrect cues) than novices. O'Hare et al. (2008) hypothesised that the superior performance of expert pilots on decision making tasks in comparison with novices occurs as a direct result of their extensive domain knowledge and the number of situation-specific schemas that experienced individuals have available to them to draw from when making decisions. Novice pilots do not have the same volume of domain knowledge and repertoire of situation specific schemas available to them.

Research by Gibb and Olson (2008) into the classification of historical United States of America Air Force aviation accidents that were determined to be as a result of human factors, found that lack of experience was a leading factor. The researchers suggested that a lack of pilot experience resulted in a significant increase in the number of perceptual and judgement errors experienced by less experienced pilots. Within the field of aviation, it is the pilot's ability to effectively make decisions based on the available environmental cues and evaluate situations under high stress and time pressure to arrive at the optimal conclusion that is vital. Experienced pilots are able to draw on their prior knowledge and training and solve problems based on "an intuitive perception of the situation" (Adams, 1993, p. 1). Both Adams (1993) and Wickens and Flach (1988) state that the biggest difference in performance between novices and more experienced pilots is the difference in knowledge stored in long-term memory that is able to be accessed on demand.

As illustrated above, the research regarding the effects of experience on the susceptibility to cognitive biases such as confirmation bias are conflicting. Much of the research showing that experience increases the likelihood that an individual will experience confirmation bias lacks ecological validity, as the experts utilised in many studies are not true domain experts. As suggested by Ericsson and Charness (1994), people must have engaged in extensive amounts of daily practice over the course of at least a decade, to be considered an expert in their field. Research by Bilalic et al. (2008a) clearly illustrates that individuals at the highest levels of expertise in chess are not susceptible to the cognitive biases that the researchers other 'expert' participants fell victim to. In addition, specifically in the field of aviation, research has clearly illustrated that experience is highly advantageous in allowing pilots to engage in quick and accurate decision-making and more experienced pilots are less likely to be involved in aircraft accidents or incidents.

2.5 Confirmation Bias in Human Error-Related Incidents

Confirmation bias in the hypothesis-testing tradition has been researched in a variety of fields and there is little doubt that the bias has a significant impact on an individual's perception and judgement. Since Wason's early research in 1960, confirmation bias has been researched extensively in the field of medicine, in the hope of finding a way to reduce errors occurring in diagnosis and treatment, and the field of criminal justice, in an attempt to identify methods for avoiding wrongful convictions (Norman & Eva, 2010; Rassin, 2010; Rassin, Eerland, & Kuijpers, 2010). Research on physicians in clinical scenarios has illustrated that the effects of confirmation bias can extend far beyond the simple process of discounting information that contradicts an individual's initial hypothesis. Research by Tschan et al. (2009) illustrated that when given a hypothesis regarding a certain diagnosis, this original diagnosis served to influence the clinical participants' auditory perceptions of the patient's symptoms and even facilitated complete auditory hallucinations in some participants.

More recently, confirmation bias research has also begun to focus on the role of confirmation bias in other human error-related incidents that have occurred outside of the fields of medicine and criminal justice. The initial difficulty with assessing the actual impact of confirmation bias in real-life situations is that it arises from a testing strategy, which under normal circumstances is quite efficient and adaptive. McKenzie (2003) also suggests that an individual's behaviour on a laboratory-based rule discovery task may seem irrational but in real-world scenarios, their strategies utilised may in fact be useful and effective. It has since been identified that the real issues occur when people encounter novel situations or certain contextual factors interact causing this normally adaptive process to become entirely inappropriate, which then results in a dysfunctional outcome (Rudolph, Morrison, and Carroll, 2009). An illustration of a situation where this occurred is the Ladbroke Grove train crash.

Lawton and Ward (2005) carried out a systematic analysis of the systems and evidence involved in the Ladbroke Grove train crash in 1999 that killed 31 people and injured more than 400 others when a Turbo train collided with a high-speed train in the vicinity of Paddington train station in London, England. Lawton and Ward (2005) suggest that the sunlight conditions on the day are likely to have produced glare, which would have made the proceed signal difficult to distinguish between yellow and red. As the train driver, Mr Hodder, was expecting to see a yellow signal, the investigators argued that it is feasible that he experienced confirmation bias and that he misperceived the red signal to stop as a yellow signal, and as a result, he continued along his current route (Lawton & Ward, 2005). Similarly, Macrae's (2009) research on the human factors-related causes of collisions in maritime accidents showed that 33.3% of accidents and incidents that occurred as a result of an erroneous interpretation of information or evidence were directly attributable to confirmation bias.

Lawton and Ward's (2005) system analysis of the Ladbroke Grove train accident highlights the complex nature of human error and the tragic consequences of confirmation bias occurring in a real-life situation. Research by Mynatt et al. (1977), Mynatt, Doherty, and Tweney (1978), Klahr and Dunbar (1988), and Teasley et al. (1994) has illustrated the realistic applications of hypothesis testing in the rule discovery task. Unfortunately, subsequent research has been unable to clearly identify any practical interventions that significantly reduce or eliminate the effects of confirmation bias from the hypothesis-testing process. Little research has also been completed that has systematically explored the factors that make an individual more susceptible to confirmation bias. McKenzie (2006) argues that confirmation bias has been widely over-represented in the literature to date and in fact, the situations in which confirmation bias actually represent a serious concern are relatively rare. One of the fields in which confirmation bias can be cause for serious concern is the field of aviation, where an error in perception, judgement, or decision-making, could result in a large loss of life and injury and extensive environmental and property damage. In the field of aviation, research by Adams (1993), Madhavan and Lacson (2006), O'Hare (2008) and Gibb and Olson (2008) has identified that more experienced pilots perform better than novice pilots on decision-making tasks and has identified that more experienced pilots are less prone to decision-making biases.

Chapter Three

Human Factors and Confirmation Bias in the Field of Aviation

3.1 Aviation Human Factors

Fatal air crashes are incredibly traumatic, as they often involve large loss of life and have severe and long lasting effects on society, the environment, the individuals on-board, and their families. As technology and aircraft design have improved, there has been a significant reduction in the number of accidents and incidents caused by mechanical error and technology failures. With this reduction, the focus has shifted to the realm of human error and the vital role that these errors play in aircraft accidents (Wiegmann & Shappell, 2003). Data indicate that human error is responsible for approximately 80% of all aviation accidents (Bazargan & Guzhva, 2011; Wiegmann & Shappell, 2003) and of those accidents resulting in fatalities, over 50% are directly attributable to errors made in the decision-making process (Wickens & Flach, 1988).

As in other transportation systems, the cause of aviation accidents and incidents cannot be attributed to a single instance of human error (Lawton & Ward, 2004; Shappell & Wiegmann, 1997). An accident caused by human error is often as a result of a complex relationship and interaction between multiple factors and variables present at the time (Madhavan & Lacson, 2006; McGregor, 2006; O'Hare, 2006; Wiegmann & Shappell, 2001). Therefore, to understand the sequence of events and the errors that occurred, it is necessary to look at all of the available evidence, relationships, and interactions present leading up to and during the occurrence. Many factors can contribute to the incidence and appearance of error. Often there are a range of internal and external factors that can draw on the attentional resources of the flight crew, resulting in less attention being paid to their surroundings. Inclement weather, unfamiliarity with equipment and aircraft type, becoming lost, outside or personal stresses, illness or injury on board, new operating procedures, lack of task knowledge and lack of experience, are just a few examples of factors that could contribute to an increase in cognitive load for the pilot, serving to distract them from their tasks (O'Hare, 2006).

Often, many factors need to be present for an accident to occur, but they must occur at the same time or in a certain sequence with other contextual variables and factors. Reason's (1990) model of human error suggests that for an accident or incident to occur, there must be an interaction between a latent and active failure. Latent (underlying) failures already exist and can lead to potential flaws in a system, e.g., flawed aircraft design, out-dated navigation maps etc. When certain events, referred to as active failures, occur, they act as triggers of the latent failures, e.g., bad weather, poor decision-making. The latent and active failures interact and as a consequence, accidents occur (Gibb & Olson, 2008). Airline cockpits are an example of an extremely complex system that is highly susceptible to human errors. Cockpit operations require a smooth, efficient, and highly skilled user interaction between human and machine, involving inputs from the Captain, the co-pilot, the flight mechanic, and the autopilot (Blomberg, 2011).

Generally, this system is well protected by strict rules, policies, procedures, and warning structures monitoring the aircraft's systems. Unfortunately, latent failures can go unnoticed, their effects can be long lasting, and their consequences delayed (ICAO, 1993). The latent failures that are crucial to Reason's (1990) model are a consequence of an action taken, as the result of a decision made, well before a situation or incident occurs. Latent failures can occur within any system of flight and can transpire in isolation, with no immediate adverse effects on flight safety. It is when these dormant latent failures are triggered, and interact, that they open up an opportunity for the pilot to make an error, which results in an active failure that overrides all of the system's defences and an accident occurs (ICAO, 1993). This process is vividly illustrated in both the Air New Zealand Flight TE901 and Comair Flight 5191 tragedies.

3.2 Confirmation Bias in Aviation

There are multiple underlying causes of human error (Gibb & Olson, 2008; Shappell & Wiegmann, 1997). In aviation accidents and incidents, the individual pilot is often the last barrier in the defence system to stop the latent failures from being triggered. As the last barrier, the ability of a pilot to make good and effective decisions in situations under stress is a vital requirement of their occupation, as the impact of even one poor decision could be disastrous. Confirmation bias is one type of human error, which can result in tragic errors in aviation decision-making. It has been attributed as a causal factor in several fatal accidents and incidents in the transportation arena (Croft, 2007; Lawton & Ward, 2004; Mahon, 1984). Pilots can be susceptible to confirmation bias through their preference to seek only information that confirms their current hypothesis and their inclination to under-weight or completely avoid information that is contradictory (Nickerson, 1998; Rassin, 2008). Hypotheses that pilots may have to evaluate could relate to scenarios involving mechanical failures, inclement weather patterns, and situations where the pilot is no longer sure of their location. For example, when a pilot is lost and disoriented and they are attempting to establish their location, general aviation lost procedures call for pilots to hypothesise a location where they think they are and then look for features on the ground to confirm their hypothesis. Confirmation bias can occur when pilots focus only on the features that are consistent with their original hypothesis and place little value on any terrain characteristics that are inconsistent (Wickens & Flach, 1988). This is illustrated in research by Gilbey and Hill (2009) who investigated the occurrence of biases in reasoning in aviation navigation.

Based on their understanding of confirmation bias and earlier aviation incident reports, Gilbey and Hill (2009) hypothesised that on a map-based location discovery task, participants would primarily utilise a hypothesis support-seeking strategy when reasoning about their location. Five different studies were conducted exploring the rates of task success in groups of student pilots, cognitive psychology students, and experienced orienteers. Three of the studies explored the efficacy of providing participants with varying levels of information on the negative effects of confirmation bias. Results indicated that overall, both pilots and cognitive psychology students used a hypothesis-confirming approach to the task and the effect of the simple intervention of providing information to the participant groups about confirmation bias was not significant (Gilbey & Hill, 2009). Interestingly, the group of experienced orienteers primarily utilised a disconfirmatory approach on the task. The student pilots in Gilbey and Hill's (2009) research were considered novice pilots and the researchers identified that future research would benefit from the exploration of experienced pilots and an investigation of the factors underlying the decreased levels of susceptibility to confirmation bias in orienteers.

Also recognising the importance of underlying factors, Madhavan and Lacson (2006) evaluated existing research into the identified psychological factors underpinning the decision-making strategies of pilots in poor weather situations and consolidated the data into a model of pilot decision-making strategies. The ability of a pilot to make effective decisions was significantly influenced by multiple factors all requiring simultaneous action and thereby significantly draining the psychological resources of the pilot. The researchers' model suggests that it is during the initial stages of the decision-making process when the pilots have to observe and acquire information under stress, where they are typically affected by errors such as confirmation bias (Madhavan & Lacson, 2006). The tragedies then occur when the pilots commit to their plan of action, which has been founded on erroneous evidence and a flawed decision-making process. This process is illustrated in a serious helicopter incident where the pilots diagnosed a failure of the high-speed driveshaft in the number one engine after a loud noise from the left engine compartment was experienced almost simultaneously with a yaw of the helicopter, and an apparent loss of power in the number one engine. In fact, the loud noise, un-commanded yaw, and apparent power loss were caused by two separate failures that occurred at almost the same time (Comer, 2009). The loud bang was caused by a duct coming loose from the environmental control system and a severe high side failure in the digital engine control unit of the number two engine resulted in the yaw and apparent loss of power in the number one engine. In this case, the pilots' initial evaluation of the available evidence indicated a driveshaft failure and the pilots became fixated on this hypothesis, subsequently missing the correct diagnosis.

> It's hard to explain why two pilots missed the engine indications. Human factors experts call it the strength of an idea, or confirmation bias (Comer, 2009, p. 23).

The pilots' original hypothesis was in fact incorrect and the action plan carried out was therefore far from optimal. According to Comer, one of the pilots felt that the enormous pressure that was experienced in the emergency resulted in a misinterpretation of the cockpit indications. The Erebus Disaster is another example illustrating the occurrence of confirmation bias in pilots.

3.3 Air New Zealand Flight TE901 – The Erebus Disaster

In 1977, New Zealand's national air carrier, Air New Zealand, began scenic flights to Antarctica. Over a two-year period, Air New Zealand conducted thirteen of these scenic flights successfully and without incident, until, on November the 28th, 1979, Air New

Zealand Flight TE901 lost radio communication with New Zealand after flying into the face of Mt Erebus, killing all 257 people on board. At the time, this was the fourth worst aviation accident of all time, and to this day, it remains the worst aviation disaster ever to occur for New Zealand. The initial investigation attributed the cause of the accident to ‘pilot error’ (Mahon, 1984; Vette, 1983). This conclusion warranted further investigation and Peter Mahon was appointed by the New Zealand Royal Commissioner to further explore the cause of the fatal Erebus Disaster. An investigation into the events leading up to the accident identified a long string of mistakes and errors, some of which had occurred over a year before Flight TE901 even left the ground. In Mahon’s (1984) investigation, he identified several mistakes and errors that each directly contributed to the Erebus Disaster.

1. In 1978, the rapid advancement in technology called for Air New Zealand to upgrade their systems. Flight plans, routes, and navigational waypoints all became digitised. When the airline was computerising their flight plans, the Chief navigator made multiple errors when inputting the coordinates that were to become the standard Antarctica flight plan. These erroneous coordinates differed from the correct figures by several degrees, putting the aircraft’s flight path twenty-seven nautical miles to the west of the original path. Despite numerous checking procedures, the discrepancy was missed.
2. All flight crew were supplied with a navigational chart showing the flight track. Despite being utilised for over a year, no one noticed that this printed version of the flight track differed from the computerised route. This chart was replaced by another one in 1979, which also did not match the correct flight path.
3. Early in November of 1979, the airline decided to alter the destination waypoint for the Antarctica flight. The change was supposed to make a track difference of only 2.1 nautical miles from the original track. Because the original track was never entered into the system correctly, the difference in track was again over 27 nautical miles. This was identified by Captain Simpson, an Air New Zealand pilot who often flew the Antarctica flight, and the information was passed on to Flight Operations with the recommendation that flight crews should be notified, but they were not.
4. The flight path was altered in the airline’s ground computer the night before flight TE901 departed for Antarctica and no members of the flight crew were notified.

In total, a senior law lecturer at Auckland University identified fifty-four navigational errors made by Flight Operations (Vette, 1983). The pilots were expecting a flight path that would take them directly up McMurdo Sound. With the change in coordinates, the flight path now directed them on a route taking them up the centre of Lewis Bay and directly over Mt Erebus. On the day of the accident, the sky was clear and the weather was perfect (Mahon, 1984). Features that the pilots expected to see in the clear air to the left and right of the aircraft, with a flight path taking them down the centre of McMurdo Sound were confirmed. The features they saw on their actual flight path approaching Ross Island were very similar to the features that their erroneous mental set led them to expect to see. The headlands visible to the left and right of Ross Island would have looked exactly like the headlands visible on the left and the right when flying into McMurdo Sound.

> The expectation of where they were flying and their understanding of where they thought they were, would have been reinforced by both the knowledge and approval given to their intended actions by the ATC McMurdo, and the topographical sightings which they made whilst visually sweeping the area prior to descent….The visual cues encountered while flying the nav track towards Ross Island were consistent with what they were expecting to see, and would have in fact seen if flying up McMurdo Sound (Vette, 1983, p. 230).

Fig. 1 clearly illustrates the path that the pilots thought that they were on and Fig. 2 illustrates the flight path that they had actually taken (Mahon, 1984, pp. 90-91). Mahon (1984) also suggests that the pilots were probably affected by a whiteout phenomenon. The whiteout phenomenon can occur as a result of flying under a pale overcast sky, over snow-covered terrain, with the light behind an aircraft, and it has the effect of perceptually flattening out the terrain so that all hills and mountains can effectively disappear. The combination of the whiteout phenomenon and the confirming terrain features all served to support the pilots' assumptions about their current location. The evidence supporting their assumptions about their location were perceived as being so strong that the five experienced members of the flight crew in the cockpit had no chance of detecting the face of Mt Erebus until it was too late.

The Erebus disaster is an important illustration of how errors can result as a consequence of a complex interaction of latent and active system failures as suggested by

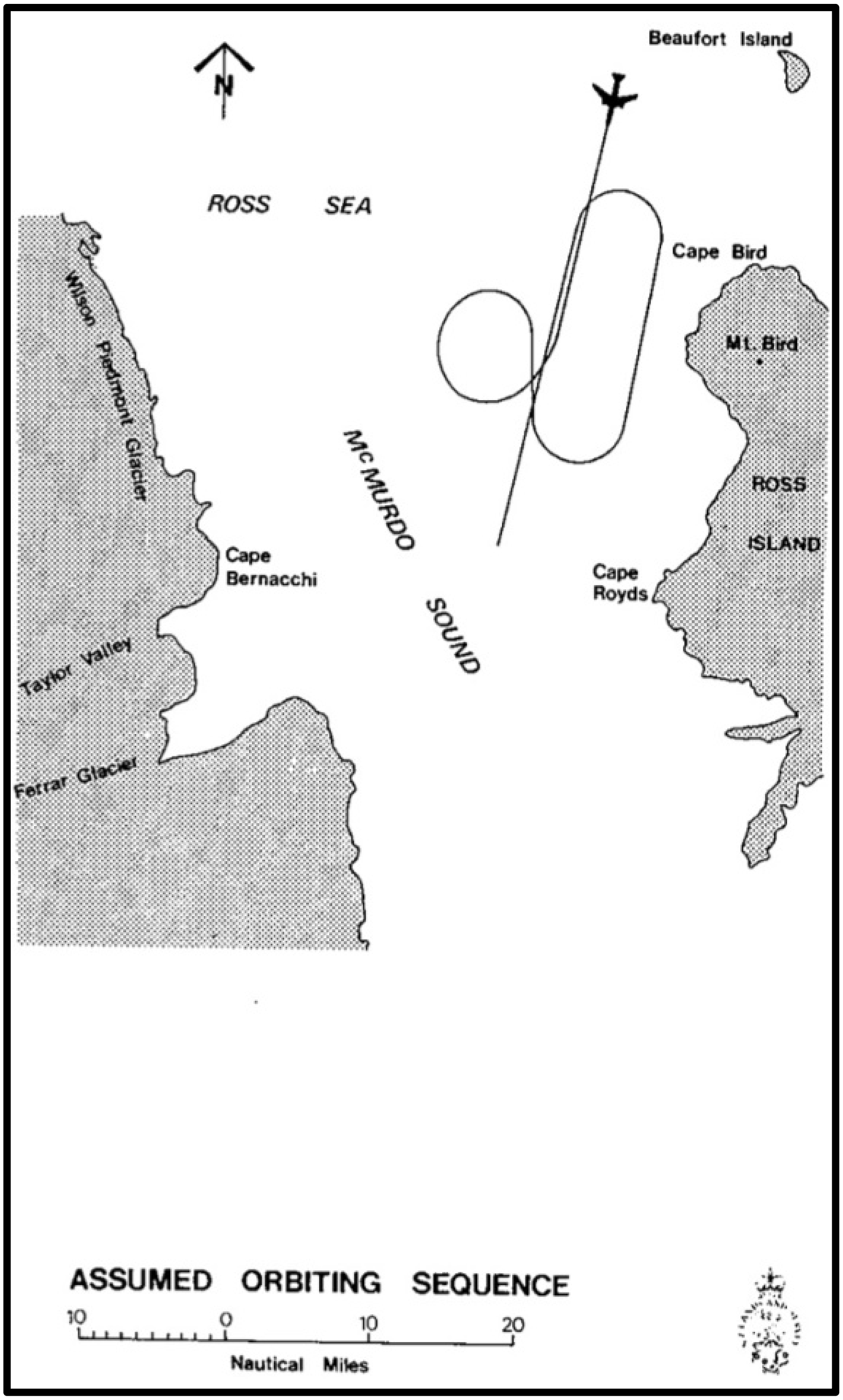


Figure 1. Assumed orbiting sequence of Flight TE901. Reproduced from Mahon (1984, p. 90).

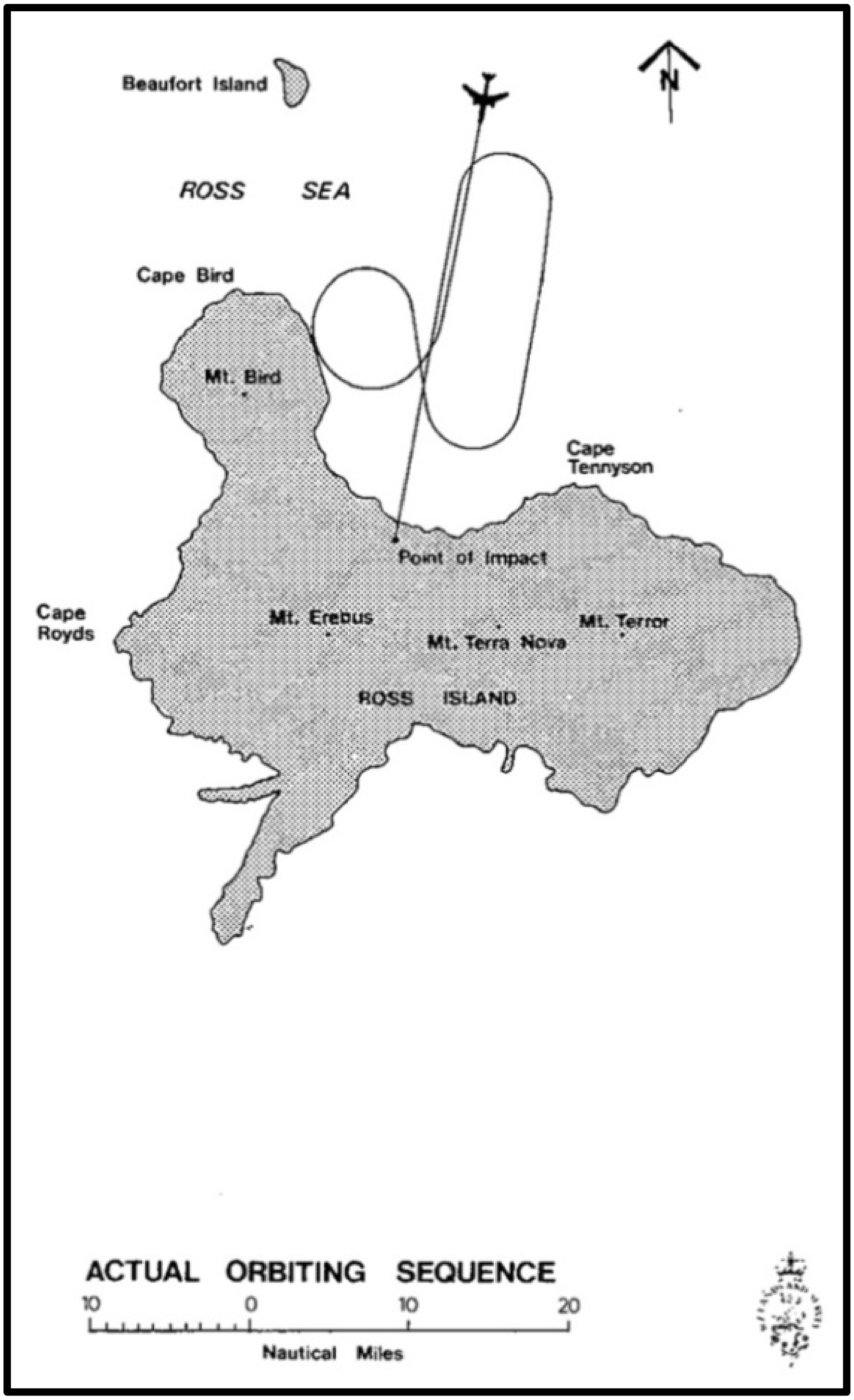


Figure 2. Actual orbiting sequence of flight TE901. Reproduced from Mahon (1984, p. 91).

Reason (1990). The interaction of multiple navigational errors made on the ground by Flight Operations staff, namely the change in flight path for which the flight crew were unaware, resulted in Flight TE901 flying not into McMurdo Sound as the pilots thought, but directly into Lewis Bay and into the face of Erebus. The complex interaction of navigation errors, whiteout phenomenon and the possible incidence of confirmation bias in the flight crew, resulted in a data entry error that occurred over a year ago turning into a disastrous accident (Mahon, 1984). "There the elements of nature had so combined, at a fatal coincidence of time and place, to translate an administrative blunder in Auckland into an awesome disaster in Antarctica" (Mahon, 1984, p. 295). The fatal effects of confirmation bias in the field of aviation are further illustrated by the Comair Flight 5191 accident.

3.4 Comair Flight 5191

On August 27, 2006, Comair Flight 5191, a fifty passenger Bombardier Canadair Regional Jet (CRJ-100), crashed during take-off at the Lexington, Kentucky, Blue Grass Airport after the pilots had attempted to take off from the incorrect runway. The crew were instructed to take off from Runway number 22 and despite having confirmed this instruction to the tower the flight crew proceeded to line up on Runway 26 and proceed with take-off. Despite the evidence of obvious factors that should have alerted the pilots to their mistake before take-off, it was believed that the pilots misinterpreted the disconfirming evidence in order for it to conform to their original hypothesis (Croft, 2007). Consequently, the aircraft failed to take off as runway 26 was far too short for a take-off of an aircraft of that size and type. The aircraft crashed into trees and other obstacles at the end of the runway killing 49 of the 50 individuals on board. The only person to survive was the first officer who received serious injuries. The airplane was destroyed by the impact and subsequent fire. Similar to Air New Zealand Flight TE901, the investigation into the cause of the accident identified almost thirty contributing factors. Two of the findings in the official accident investigation report indicate the occurrence of confirmation bias (National Transportation Safety Board, 2007):

1. "Adequate cues existed on the airport surface and available resources were present in the cockpit to allow the flight crew to successfully navigate from the air carrier ramp to the runway 22 threshold" (p. 103).
2. "The flight crewmembers failed to recognize that they were initiating a

> take-off on the wrong runway. They did not crosscheck and confirm the airplane's position on the runway before take-off and as a result, they were likely influenced by confirmation bias" (p. 103).

Transcripts provided from the cockpit voice recorder indicated that the pilots were expecting the runway end identifier lights on runway 22 to be out of order, therefore when the pilots made the error of turning on to runway 26, it would not have seemed strange to the flight crew for these lights to be absent. The investigators concluded that the flight crew evaluated the visual cues and evidence surrounding them on the ground at the airport in a confirmatory fashion, allowing them to confirm their hypothesis that they were already at the threshold for runway 22 despite the abundance of disconfirmatory cues available to the flight crew, including:

> The runway holding position for runway 26, the 75-foot painted width of runway 26 (versus the 150-foot width of runway 22), and the absence of runway edge lights and precision runway markings (such as threshold markings and touchdown zone markings) on runway 26 (National Transportation Safety Board, 2007, p. 66).

All of the information available to the crew that would disconfirm their hypothesis was ignored and the pilots proceeded to pause at the threshold for runway 26 before turning on to it and attempting to take off. It was not until the pilots were already well into their take-off roll that the mistake was identified and by this stage it was too late to abort take-off (National Transportation Safety Board, 2007). The pilots attempted to bring the plane into the air, but there was not enough power and the plane consequently crashed into terrain at the end of the runway.

3.5 Current Research Aims and Hypotheses

Early research by Wason (1960) clearly identified a tendency for participants to primarily utilise hypothesis-confirming information on his rule discovery task. Wason termed this strategy an enumerative strategy and Wetherick (1962) was quick to criticise Wason's method, identifying that the design of Wason's RDT meant that it was impossible to accurately identify exactly what strategy participants were truly using. Further research took Wetherick's criticisms into consideration, providing alternative explanations of Wason's original ideas (Farris & Revlin, 1989a, 1989b; Klayman & Ha, 1987, 1989). Mynatt et al. (1977, 1978) coined the term *confirmation bias* and pioneered research into the investigation of confirmation bias in real-world scenarios. Klahr and Dunbar (1988) and Teasley et al. (1994) followed on from Mynatt et al. (1977, 1978) expanding on the available research into realistic applications of hypothesis testing in the RDT. The research exploring confirmation bias in real-world scenarios illustrated that confirmation bias is just as pervasive in situations occurring outside of the laboratory. More recently, research by Gilbey and Hill (2009) illustrated that student pilots primarily utilised a hypothesis support-seeking strategy on a map-based location discovery task. Following on from this prior research, *it is hypothesised that individuals will be more likely to utilise a hypothesis-confirming strategy on an aviation navigation based location discovery task.*

Despite the disastrous consequences of confirmation bias occurring in pilots, as illustrated in both the Erebus and Comair disasters, there has been minimal research that has systematically looked at the constructs that influence susceptibility to confirmation bias in pilots. The majority of current research has focused on error taxonomies and not on the reasons that these errors occur, or what aspects make an individual more susceptible to experiencing certain errors. Research exploring the effects of experience on the susceptibility to confirmation bias has been conflicting. Research by Wiley (1988), Hecht and Proffitt (1995), Lewandowsky and Kirsner (2000) and Bilalic et al. (2008a, 2008b, 2010) indicate that the more experience a participant has in their field, the more likely they are to be vulnerable to cognitive biases such as confirmation bias, especially on tasks involving novel or unfamiliar task stimuli. By contrast, research specifically conducted in the area of aviation by Wickens and Flach (1988), Adams (1993), Madhavan and Lacson (2006), O'Hare et al. (2008), and Gibb and Olson (2008), has illustrated that more experienced pilots perform better than novice pilots on decision-making tasks and experienced pilots are less susceptible to cognitive biases. Therefore, *it is hypothesised that experienced pilots will be less susceptible to confirmation bias than novice pilots, who in turn will be less susceptible to confirmation bias than non-pilots.*

Chapter Four

Method

The present research was judged to be of a low risk to participants. A low risk notification was recorded on the Massey University 'Low Risk Database', which is reported in the annual report of the Massey University Human Ethics Committees. During the completion of the research, consideration was made at all times to ensure the safety, wellbeing, and privacy of all participants and to ensure that the research followed the Massey University Code of Ethics.

4.1 Participants

In total, 53 participants took part in the research study. Of the 53 participants, 34 (64.2%) were males and 19 (35.8%) were females. The age of the participants ranged between 18 and 70 with an average age in years of M = 32.2 (SD = 13.2). The number of logged flying hours as indicated by the participants, ranged between 0-15000, M = 977.4 (SD = 2721.3). Participants were recruited through an email advertisement disseminated through the psychology graduate email distribution list at Massey University in New Zealand. A further email invitation was also distributed through the Massey University School of Aviation (see Appendix B). The target participants were pilots with any amount of logged flight experience in an aircraft. Section 2 of the Civil Aviation Act 1990 operationally defines the term 'aircraft' as follows: "Aircraft means any machine that can derive support in the atmosphere from the reactions of the air otherwise than by the reactions of the air against the surface of the earth" (Civil Aviation Authority, 1990, p. 13). Individuals were encouraged to participate if they had logged flight experience in any of the following categories: un-powered fixed wing (glider), powered fixed-wing (aeroplane) or powered rotary-wing (helicopter).

Participants were assigned to a group based on the number of logged flying hours they had completed in an aircraft. For the purposes of this research study, a novice pilot was operationally defined as an individual that had less than 200 hours of logged flying time and had not yet gained their commercial pilot's license. An experienced pilot was operationally defined as an individual who had a minimum of 201 logged flying hours. These guidelines were selected based on the Civil Aviation Authority in New Zealand's minimum requirement of 200 hours logged flight time, for an individual to be eligible for a Commercial Pilot's License (Civil Aviation Authority, 2011). Therefore, individuals with between 1 and 200 logged flying hours were assigned to the novice group, individuals with more than 200 logged flying hours were assigned to the experienced pilots group, and those participants with zero logged flying hours were assigned to the control group (non-pilots).

Of the 53 participants, there were 23 participants in the control group (43.4%), 13 in the novice pilot group (24.5%) and 17 in the experienced pilot group (32.1%). The average age of participants in the control group was *M* = 32.8 (*SD* = 10.7) with a range of 22-65 and there were *n* = 7 males (30.4%) and *n* = 16 females (69.6%). In the novice pilot group the average age was *M* = 32.9 (*SD* = 19.3) with a range of 18-70 and there were *n* = 12 males (92.3%) and *n* = 1 female (7.7%). The average age of participants in the experienced pilot group was *M* = 30.8 (*SD* = 11.1) with a range of 20-67 and there were *n* = 15 males (88.2%) and *n* = 2 females (11.8%). There was a marked difference in the distribution of gender across the three groups. The low rates of female participants in both the novice and experienced pilot participant samples is in line with current Federal Aviation Administration data, indicating that female pilots in the United States of America account for only 6% of currently licensed pilots (Federal Aviation Administration, 2011). The mean number of logged flying hours for the novice pilot group was *M* = 78.5 (*SD* = 75.1) with a range of 12-200. The mean number of logged flying hours for the experienced pilot group was *M* = 2987.1 (*SD* = 4212.3) with a range of 220-15000.

4.2 Materials and Stimuli

The present research followed a between subjects experimental design that aimed to investigate the role of experience in the susceptibility to confirmation bias in non-pilots, novice pilots and experienced pilots, and to explore the occurrence of any different decision-making strategies between these three groups. In order to achieve this, the present investigation utilised four map-based scenarios that represented four location discovery tasks. The experimental design and procedure was based on a realistic application of the theory behind Wason's (1960, 1968) original rule discovery tasks and closely followed research conducted by Gilbey and Hill (2009) who investigated the incidence of confirmation bias in aviation navigation procedures. The independent variable – level of experience, was investigated on three levels, control (non-pilots), novice pilots and experienced pilots. The dependent variable – susceptibility to confirmation bias, was measured by a participant's success on the four location discovery tasks and was represented by a total score out of four. In the construction of the present research, recommendations of previous research into the limitations of investigations in the hypothesis testing tradition to date were taken into account.

Success on research tasks involving rule discovery or hypothesis testing is greatly increased through the use of familiar features and scenarios that participants are able to relate to (Gale & Ball, 2009; McKenzie, 2006; Van der Henst, Rossi, & Schroyens, 2002). Utilising realistic material instead of abstract features also fosters increased rates of success (Jones & Sugden, 2001; Wason & Shapiro, 1971). Therefore, because the current research is interested in the susceptibility to confirmation bias in pilots, the research looks at reasoning in an aviation context to ensure that the content is familiar and relevant to the participants' experience. Simply having thematic content as opposed to abstract constructs as part of the research design is not sufficient; the thematic content must also be pertinent to the target participants' background and experience (Reich & Ruth, 1982). In line with these previous research indications, the four location discovery tasks utilised in this research were designed to be as ecologically valid as possible for the target participants. As the present research was primarily interested in the role of experience in the susceptibility to confirmation bias in pilots, all of the scenario tasks followed a primary aviation theme. Real life scenarios were created describing tasks and situations that a pilot flying under visual flight rules (VFR) could potentially expect to encounter during flight. The use of current New Zealand aviation visual navigation charts also aided in making the scenarios as realistic as possible. The visual navigation charts utilised were issued in 2010 and were approved for use in general aviation by the Civil Aviation Authority in New Zealand. The scenarios developed were constructed from visual navigation charts B5 (Central Plateau) and B6 (Southern Alps).

Utilising these New Zealand aviation visual navigation charts and realistic aviation themes, four location discovery scenarios were developed. See Appendix C for the scenarios used in the location discovery task. Instead of reasoning about a rule, the participants had to engage in a hypothesis-testing task involving reasoning about their location in an aviation navigation task where they were unsure of their current location. General aviation lost procedure instructs pilots to 1) orbit a feature 2) draw a circle of uncertainty (COU) on their map and 3) look for a distinctive feature on the ground and try to find it in their COU. The participants' hypothesised location represented by the COU was provided to them. This hypothesis was identified as R_H representing the hypothesised (H) rule (R). For all of the scenarios the participants were told that they were in fact located elsewhere on the map in an area similar to their hypothesised location, but differing in one distinct way. This location is referred to as R_T or the true rule. See Appendix D for the aviation visual navigation charts utilised in the current scenarios illustrating R_H and R_T. Just like Wason's (1960) research, which provides number triples to participants to enable them to find a numerical rule, the current research design utilised three features to represent the triples from Wason's original study (Gilbey & Hill, 2009). Two out of the three features corresponded to the participants' hypothesised location and all three features corresponded with their true location. The four scenarios were then uploaded on to the Massey University experiment server so that the research could be easily accessed by participants from anywhere in New Zealand. To control for order effects a balanced Latin Square design was used to counterbalance the ordering of the four scenarios. It was necessary to follow this method to ensure that participants' results on the task were not being adversely impacted by the order in which the scenarios were given. The four scenarios were presented in the following manner: Order 1) 1, 2, 4, 3, Order 2) 2, 3, 1, 4, Order 3) 3, 4, 2, 1, Order 4) 4, 1, 3, 2.

4.3 Procedure

Participants were able to access the experiment from any computer that had an internet connection by following the web link provided on the research invitation. The web link directed prospective participants to the information sheet for the research and after reading the information sheet, participants could then consent to participating by clicking a hyperlink through to the research experiment. The ordering of the scenarios was an automatic process and every time a participant completed the scenario in one order condition, the next participant would be given the scenarios in the second order condition. This process continued for each participant cycling through the available order conditions.

After participants had been given the chance to read the general instructions and consent to the research, information about their gender, age, and flight experience was gained before they were directed to the first scenario. Following the procedure of Gilbey and Hill (2009) all participants were instructed to imagine for each scenario that they were lost, but that they had constructed a hypothesis about where they were located. For each of the four scenarios, the hypothesised location was indicated on the visual navigation chart provided in the form of a circle that represented their circle of uncertainty (COU) and their probable location. Participants were then asked to make a judgement about their location on the map by identifying from a list of three distinctive features, which feature would help them to distinguish most accurately whether they were within their hypothesised location or not. After each of the four scenarios had been completed, each individual was asked to explain their reasoning and the rationale for their choice.

In order to determine the participants' susceptibility to confirmation bias their decision-making strategies on this scenario task were examined. As indicated in Gilbey and Hill (2009), the participants who chose the inconsistent feature were said to be making the 'correct' decision. Participants who chose an identifying feature consistent with both locations were considered to have engaged in hypothesis-confirming decision-making (i.e., exhibited confirmation bias). Those participants who chose as their most important distinguishing feature, the one feature that was not found in their hypothesised location were said to be utilising disconfirming evidence in their decision-making and as a result, had not succumbed to confirmation bias.

Chapter Five

Results

5.1 Data Analysis Procedure and Preliminary Analyses

All data analyses were completed utilising IBM's SPSS v19 for Windows. The total number of disconfirmatory responses out of four was calculated for each participant to determine whether individuals were primarily following a hypothesis-confirming, or a hypothesis-disconfirming technique in the location discovery task. A participant's total score out of four represented their susceptibility to confirmation bias, with a score of zero illustrating the use of a hypothesis-confirming method throughout the task and a score of four indicating that a participant had utilised a hypothesis-disconfirming approach across all of the scenarios. Performance within each of the four different scenarios across the groups was also investigated. To assess the relationship between the independent variable – experience, and the dependent variable – susceptibility to confirmation bias, a participant's total amount of flying experience in hours was correlated with their total score out of four on the location discovery task. An analysis of variance was performed to assess whether the between-groups differences in the effects of experience on the susceptibility to confirmation bias, were greater than the expected within-group differences. The comments given by each participant after each scenario were analysed and assessed to identify common decision-making strategies utilised by the three participant groups. Participants were also assessed for consistency in their decision-making strategies across the four scenarios, to identify whether or not individuals were following the same decision-making strategies throughout the location discovery task.

Preliminary analyses were carried out to check that the data was error free, assess the data for any violations of the assumptions underlying the statistical techniques to be utilised in the data analysis and check that the results had not been impacted by order effects. The full data set was checked for errors, including the identification of any missing cases and any values that fell outside the possible range of values for all variables, both categorical and continuous. Assessment of the data found no missing values and confirmed that all data fell within valid limits. Skewness and kurtosis values were also obtained for the dependent variable, susceptibility to confirmation bias. The skewness statistic of .709 indicated a positive skew in the data set with scores on the location discovery task clustered to the left at the low values. The kurtosis statistic of -.253 indicated a relatively flat distribution of scores. The difference between the mean for the total number of scenarios correct out of four and the 5% trimmed mean was .06 indicating that there were not any extreme values distorting the results. The Kolmogorov-Smirnov statistic was also obtained to test for normality and the significant result indicated a violation of the assumption of normality. To further investigate, a scatterplot, a histogram, and normal probability plots were constructed. The reasonably straight line on the normal Q-Q plot suggested a relatively normal distribution but the shape of the scatterplot and the histogram highlighted the positive skew as indicated by the earlier skewness statistic. The de-trended normal Q-Q plot also showed a deviation from the normal distribution with most points not collecting around the zero line.

The assumptions of level of measurement, related pairs, and independence of observations were met. The constructed scatterplot indicated a violation of the assumptions of normality, linearity, and homoscedasticity. As these assumptions must have been met in order to calculate Pearson's product-moment correlation coefficient (*r*), the non-parametric alternative of Spearman's rank order correlation (*rho*) was utilised in the correlation analysis. Preliminary analyses also identified that the assumption of homogeneity of variance had been met, as Levene's test for the equality of variances was not significant on both the ANOVA and t-test analyses. As statistical techniques such as the ANOVA and the t-test are robust to the violation of the assumption of normality, especially with sample sizes of over 30 (Pallant, 2011), the decision was made in the present research to continue with the use of an ANOVA in the present data analysis despite the violation of the assumption of normality. An ANOVA was also completed during the preliminary analyses to check for any existing order effects. The assessment detected no significant effect of order on the results; therefore, all subsequent analyses ignored this factor.

5.2 Performance on the Location Discovery Task

Performance within the four scenarios varied widely between scenarios one, two, and three, and scenario four. Scenario four clearly promoted the greatest levels of bias proneness, with only n = 3 (5.7%) participants selecting the correct feature. Scenarios one, two, and three appeared to promote a similar level of bias proneness, with *n* = 15 (28.3%), *n* = 20 (37.7%), and *n* = 20 (37.7%) participants respectively, scoring correctly on these three scenarios. Performance rankings also varied widely across scenarios.

1. Scenario One: 1st Non-pilots, 2nd Experienced pilots, 3rd Novice pilots.
2. Scenario Two: 1st Experienced pilots, 2nd Non-pilots, 3rd Novice pilots.
3. Scenario Three: 1st Non-pilots, 2nd Novice pilots, 3rd Experienced pilots.
4. Scenario Four: 1st Novice pilots, 2nd Experienced pilots, 3rd Non-pilots.

Despite this variation, Chi-square tests for independence between group membership and scenario response (incorrect or correct) identified no significant differences between groups with regards to the scenarios which group members were the most prone to bias on. Scenario one, χ^2 (2, *n* = 53) = .84, *p* = .66, *phi* = .13, scenario two, χ^2 (2, *n* = 53) = .98, *p* = .61, *phi* = .14, scenario three, χ^2 (2, *n* = 53) = 2.46, *p* = .29, *phi* = .22, and scenario four, χ^2 (2, *n* = 53) = .18, *p* = .92, *phi* = .06. To determine participants overall success on the location discovery task their total score out of four was calculated. Overall, success on the task was poor, with an average score out of four across all participants of *M* = 1.1, *SD* = 1.1. See Table 1 for an overall summary of participants' total scores out of four.

Table 1

Summary of Participants' Total Scores Out of Four on the Location Discovery Task

Score	Frequency	Percent	Cumulative Percent
0	19	35.8	35.8
1	17	32.1	67.9
2	11	20.8	88.7
3	5	9.4	98.1
4	1	1.9	100.0
Total	53	100.0	

As the table illustrates, 67.9% of participants scored either zero, or one out of four on the location discovery task and only one participant scored four out of four correct. These results support hypothesis one, illustrating that overall, participants primarily utilised a hypothesis-confirming strategy on the location discovery task.

Participants' scores were also examined by group to identify any initial differences between the groups on task success. The non-pilot group scored *M* = 1.2, *SD* = 1.0, the novice pilot group scored *M* = 1.0, *SD* = 1.3 and the experienced pilot group also scored *M* = 1.0, *SD* = .9. Initial analyses illustrated that participants' task performance was very similar between groups.

5.3 Effects of Experience on the Susceptibility to Confirmation Bias

The relationship between experience, measured by a participant's total number of logged flying hours, and their susceptibility to confirmation bias, measured by the total number of scenarios that participants scored correctly out of four (TNC – Total Number Correct), was investigated using Spearman's rank order correlation coefficient (*rho*) as preliminary data analyses performed indicated violations of the assumptions of normality, linearity, and homoscedasticity. No relationship between experience and a participant's susceptibility to confirmation bias was identified, *rho* = -.10, *ns.* Correlations were also calculated between TNC and gender (*rho* = .18, *ns)*, TNC and age (*rho* = .04, *ns)* and TNC and scenario order presentation (*rho* = -.21, *ns)*, to ensure that these variables were not interfering or interacting with the effect of the independent variable on the dependent variable in any way. None of these analyses were statistically significant.

A one-way between-groups analysis of variance was conducted to explore the impact of experience on participants' susceptibility to confirmation bias as measured by the location discovery task. Participants were divided into three groups according to their stated number of logged flying hours (Group 1: 0 logged flying hours; Group 2: 1 to 200 logged flying hours; Group 3: 201+ logged flying hours). There was no statistically significant difference at the $p < .05$ level in confirmation bias scores for the three experience level groups: F (2, 50) = .27, *ns.* The effect size, calculated using η^2, was .01 illustrating that the amount of total variance in a participant's susceptibility to confirmation bias that is predictable from knowledge of their levels of flying experience is extremely low.

A single sample t-test was also utilised for each comparison with the test value set at 1.3, representing the mean number of disconfirmatory responses that would be expected on the location discovery task by chance alone. The mean number of disconfirmatory choices was as follows: non-pilots *M* = 1.2, *SD* = 1.0; *t* (22) = -.38, *ns.* Novice pilots *M* = 1.0, *SD* = 1.3; *t* (12) = -.30, *ns*. Experienced pilots *M* = 1.0, *SD* = .94; *t* (16) = -1.32, *ns*. Fig. 3 illustrates the percentage of scores received overall factored by group.

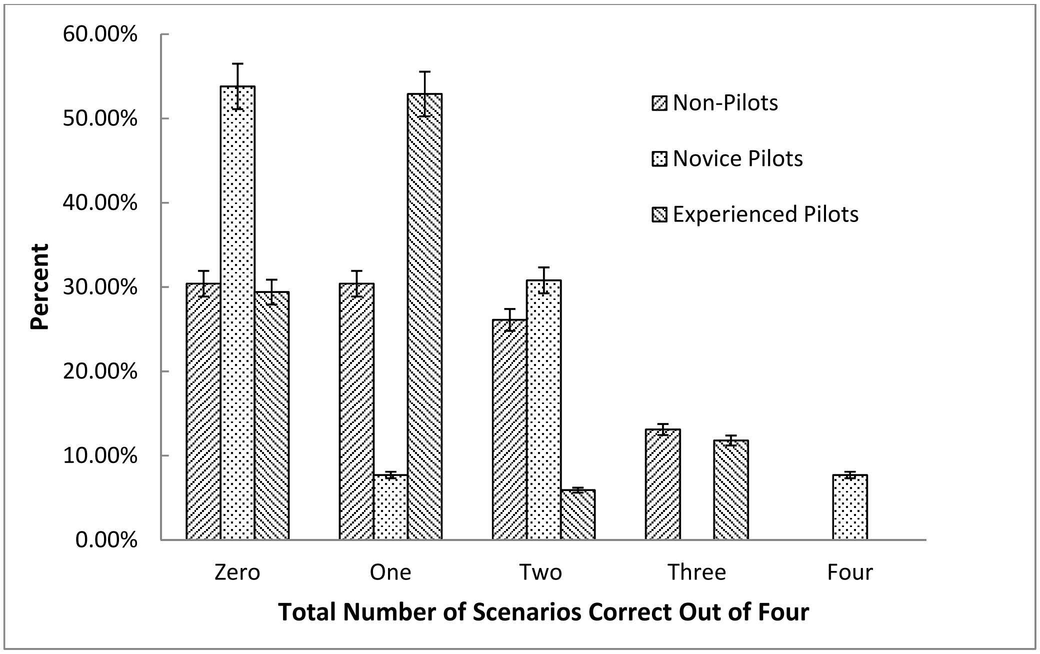


Figure 3. Summary of participants' total scores out of four factored by group.

The results obtained from the t-test and ANOVA analyses indicate that hypothesis two - participants with higher levels of flying experience will be less susceptible to confirmation bias, was not supported.

5.4 Feature Selections

In scenario one, the most widely selected feature was feature one *n* = 26 (49.1%). In scenario two, the most widely selected features were features two and three with *n* = 20 (37.7%) each. In scenario three, the most widely selected feature was feature two *n* = 20 (37.7%) and in scenario four, the most widely selected feature was feature one *n* = 40 (75.5%). See Table 2 for a summary of the most common feature selected for each scenario in comparison with the correct feature choice.

Table 2

Summary of the Most Common Feature Selected in Comparison with the Correct Choice

Scenario	Correct Feature	Correct	Most Common Choice	Percentage
1	Large pine forestry area to the east, west, and north.	28.3%	Town directly below the aircraft.	49.1%
2	Road and Railway run side by side below the aircraft.	37.7%	Road and Railway run side by side below the aircraft.	37.7%[a]
			Tall mountain summit to the west.	37.7%[a]
3	Area of swamp and marshlands below the aircraft.	37.7%	Area of swamp and marshlands below the aircraft.	37.7%
4	Large area of pine forest to the east.	5.7%	Road and Railway run side by side below the aircraft.	75.5%

[a] In scenario two, feature two and feature three were both selected equally.

Interestingly, despite the overall poor performance of participants on the location discovery task, the correct feature was chosen as the most common choice in both scenario two and scenario three. Further analysis of participants' comments received describing the reasons behind their feature choices, indicated that approximately 25% and 50% of participants in scenarios two and three respectively, did not choose the correct feature as a result of actively seeking out hypothesis disconfirming information. In scenario three specifically, the high percentage of participants scoring correctly despite not actively seeking disconfirming information, suggests that participants may have been drawn to select this feature as a result of its uniqueness in comparison to the other given features, as it occurred in only one location on the map. For example, Participant 11: *"Towns and mountains are everywhere. Swamps are less common and hence more useful."* Despite the variances in scenario two and three, overall across the four scenarios, n = 58 correct feature selections were made out of *n* = 212 total feature selections. Of the 58 correct feature selections, *n* = 37 (64%) were correctly selected because the participant was actively seeking out information to disconfirm the current hypothesis. Overall, *n* = 44 (21%) feature selections made were based on a hypothesis disconfirming strategy and of these *n* = 37 (84%) led to the selection of the correct feature.

Further investigation of the feature selections for each scenario also identified differences between the three participant groups in scenarios two and three. There were no differences in the response patterns across the three groups within scenarios one and four, with all three groups selecting the same feature as their most common choice in both scenarios. In scenario two, both the non-pilot group and the novice pilot group selected feature three most often, and the experienced pilot group selected the correct feature, feature two, most often. Interestingly, feature three was the least popular option for the experienced pilot group, despite being the most popular feature for the non-pilot and novice pilot groups. For scenario three, all three groups selected different features most commonly with the non-pilot group selecting the correct feature, feature two, most often.

5.5 Thematic Analysis of Participants' Reasons Behind Their Feature Selections

A thematic analysis of the comments provided by participants was carried out as a subsequent analysis to identify the strategies that participants were using to reason about their location and to identify any differences in the strategies utilised between the three participant groups. The comments were analysed by the researcher alone, by reading through all of the comments provided by participants after each scenario. Through the initial reading of the feedback provided, common themes were identified that illustrated participants' different decision-making strategies. Secondly, all of the comments provided were assessed for key words relating to the identified themes, and comments were then categorised accordingly. Overall, seven common decision-making strategies were identified. These were as follows:

1. Large Features – For example: mountain ranges and mountain peaks. Participants utilising this strategy primarily focused their attention on large features first, as being the most important feature to aid them in their decision-making process.

Participant 20: *"Big to small."*

Participant 34: *"A mountain summit should be a familiar landmark and an easy one to orientate with."*

Participant 42: *"Lost procedure starts with identifying the big picture and working your way in."*

2. Manufactured Objects – For example: towns, roads, and railways. Participants utilising this strategy primarily focused on manufactured objects as being the starting point for their decision-making process.

   Participant 28: *"Mountain ranges can look similar in several locations. Forests often differ from the map due to felling of trees, the road and railway usually offer a unique location."*

   Participant 18: *"A large town is the easiest landmark to spot and also gives you again an exact location. The others will only give a rough guide to where you are."*

   Participant 31: *"By circling over the town I will be able to find at least three features to fix my position."*

3. A focus on features occurring outside the specified circle of uncertainty. Participants utilising this strategy, primarily focused all of their attention on features outside the given circle of uncertainty.

   Participant 12: *"1 and 2 are features that are below the plane and would be harder to see."*

4. A focus on features occurring inside the specified circle of uncertainty. Participants utilising this strategy, primarily focused all of their attention on only those features that occurred inside the given circle of uncertainty.

   Participant 4: *"This is the landmark that is concentrated in circled area."*

   Participant 26: *"Easier to obtain a navigational fix from an object below the aircraft."*

5. Specifically seeking out features that confirmed the hypothesised location. Participants following this strategy primarily sought out features that confirmed the given hypothesised location.

   Participant 28: *"The town has a distinctive shape confirmed by the positions of the road/railway/hills."*

Participant 33: *"by looking at how the road and the railway line interact (the direction and curves) along with their position relative to the town you are above you would be able to confirm whether the town you are above is the one you think it is. Would also need to see a road coming into the town from the southwest to confirm it was Te Kuiti."*

Participant 48: *"Town is large enough so as not to mistake for a smaller village etc and has enough of a defining shape to again confirm it is the correct town."*

6. Specifically seeking out features that disconfirmed the hypothesised location. Participants following this strategy primarily followed a hypothesis disconfirming approach in their decision-making.

Participant 13: *"There is no pine forest in the circle - the large pine forest is to the East and so I can't be in the circle."*

Participant 35: *"If you look at other areas of the map you can see there are several shaded towns and numerous roads and railways that run side by side. However there is no forestry in the circle."*

Participant 2: *"Both the buildings and mountain summit appear to the west (buildings more north-west and mountain more south-west) of the red circle. Road and railway are not below - they are south of the red circle."*

Participant 11: *"railway line runs north / south. Mountain is south east towns are everywhere. I am NOT in the circle!"*

Participant 50: *"There are no roads running alongside railway lines within the circled area."*

Participant 9: *"There is only one area of swamp shown on the chart. The other items are true for the red circle but the swamp is not true so is the decider when determining position."*

7. Static Objects – Features that participants subjectively felt were reliable as being consistent over time. Participants utilising this strategy primarily focused their attention in their decision-making process on features that they subjectively felt would be consistent over time and therefore the most important in helping them identify their true location.

Participant 1: *"Because towns look variable. Railway lines and woods don't."*

Participant 21: *"The road rail will is the only feature which is unlikely to have changed since the map was printed. By assessing where the road crosses the rail and what other features such as towns other roads it should be easy to quickly determine your position. There are several towns within the circle also forestry areas often change and so are not reliable."*

Comments that were unclassifiable as they were either intentionally blank or the participant had misunderstood the question, were coded as 99. All of the participants' comments were then coded as above. Initial results indicated that overall across the three participant groups, participants were fairly consistent with their decision making strategies, with 87% of participants utilising the same decision making strategy at least 50% of the time, 40% of participants using the same decision making strategy at least 75% of the time and 9.4% of participants used the same decision making strategy 100% of the time across all four scenarios.

Actively selecting feature options containing manufactured objects was the most popular strategy overall, accounting for 25% of reasons behind specific feature selections. Actively selecting feature options containing large features, the specific search for confirming evidence and the specific search for disconfirming evidence were all also very popular strategies with 21% of feature selections made attributable to these three strategies. The least favourable strategies were specifically focusing on features outside the circle of uncertainty (0.5%), specifically focusing on features inside the circle of uncertainty (3%), and specifically focusing on static objects (3%), combined, accounting for only 6.5% of the feature selections. Comments that were unclassifiable accounted for 6% of selections. The most popular decision making strategies utilised for each participant group were as follows: both the non-pilot group and the novice pilot group utilised confirming evidence most regularly and also favoured large features and manufactured objects, in contrast, the experienced pilot group favoured manufactured objects and large features in their feature selection. These results further support hypothesis one, illustrating that participants in the non-pilot group and the novice pilot group predominantly favoured a hypothesis-confirming strategy in their decision-making. Interestingly, despite the overall poor performance of the experienced pilot group in the location discovery task, a hypothesis-confirming *strategy* was not predominantly utilised in any of the four scenarios in contrast with the other two participant groups.

Scenario one and scenario four both included some features that were located inside and some features that were located outside the given hypothesised circle of uncertainty. In scenario one, n = 38 (72%) of participants selected features located within the boundaries of the circle and in scenario four, n = 40 (75%) of participants selected features located inside the circle. These results indicate that overall, participants were actively seeking features to confirm their hypothesised location within the given circle of uncertainty, further supporting hypothesis one.

Chapter Six

Discussion

6.1 The Occurrence of Confirmation Bias

Wason's (1960) rule discovery task and derivatives of the original task framework have been utilised frequently in research studies over the past fifty years. Wason's (1960) research findings that illustrated a tendency for participants to utilise a hypothesis-confirming strategy on the RDT, fuelled the following decades of investigation into the ways in which individuals evaluate and test hypotheses. The current study aimed to examine the occurrence of confirmation bias in non-pilots, novice pilots, and experienced pilots on an aviation based location discovery task that was based on a realistic application of the theory behind Wason's RDT. It was hypothesised that overall, individuals would be more likely to utilise a hypothesis-confirming strategy on the task than a hypothesis-disconfirming approach. The results of the present research were consistent with the results of Wason's (1960) research findings and more specifically, recent research by Gilbey and Hill (2009) examining the hypothesis-testing behaviour of both non-pilots and pilots on a location discovery task. Similar to both Wason (1960) and Gilbey and Hill (2009), overall, participants in the present study performed poorly and the research identified a strong tendency for participants to utilise a hypothesis-confirming approach on the task.

The present research also illustrated that participants, when given an opportunity to seek out and identify task features, more commonly selected features that pertained to and supported the hypothesis that was supplied in the scenario, even when one of the task features provided did not match the hypothesised location. Despite the fact that a falsification of any of the four hypothesised locations could be made with the identification of the one feature that did not correspond with the hypothesised location, participants appeared to ignore the incongruent feature completely, or place little importance on it. Individuals also exhibited a high amount of confidence in their feature selection and attempted to find and provide reasons to support the given hypothesis and rationalise their feature selection. Little mention was made of the disconfirming feature except in comments made by participants who were predominantly following a hypothesis-disconfirming strategy. These individuals were more likely to critically evaluate each of the task features and quickly identify the task feature that did not fit with the given hypothesis. In the present study, a hypothesis-disconfirming method of hypothesis evaluation was the most powerful strategy to use, with over 80% of feature selections made utilising this approach resulting in the correct answer.

Generally, participants who scored correctly on each scenario, regardless of their group membership, appeared to do so as a direct result of seeking out hypothesis-disconfirming information and not by chance, with over 60% of the correct feature selections made overall, being made as a result of participants actively seeking out hypothesis-disconfirming information. Scenario three was an exception to this conclusion, with 50% of participants scoring correctly despite not following a hypothesis-disconfirming approach. It is predicted that this occurred because one of the features available for selection (area of marsh) was unique and only occurred once on the map provided for this scenario. Spellman, López, and Smith (1999) looked at the consistency of their participants' scores across their four research scenarios investigating how individuals utilise available evidence to narrow down their hypotheses. Across their three experiments, strategy consistency among participants was 81%, 75%, and 67% respectively, illustrating that individuals are relatively consistent in their decision-making strategies. Similarly, participants in the present study were also relatively consistent in their decision making approach with almost 90% of participants following the same strategy for at least two out of the four scenarios. Strategy consistency was influenced by the availability of a unique feature, as many participants altered their hypothesis-testing approach when a unique feature, such as the area of marsh in scenario three, was available as a selection choice. It is speculated that the uniqueness of the incongruent feature in this scenario may account for the change in hypothesis-testing strategy illustrated by participants. It is predicted that when the incongruent feature is also unique to the scenario, it can result in participants changing their hypothesis-testing strategy and scoring correctly despite not using a specific hypothesis-disconfirming approach.

Overall, individuals did not appear to be systematically searching through the available features to confirm their placement on the navigational charts before making a decision. The thematic analysis illustrated that participants appeared to be evaluating the usefulness of each feature in aiding them to decide whether they were in the hypothesised location, based on their prior knowledge of the features contained in each scenario and may not have been utilising the map at all in their decision making process. This phenomenon was also illustrated by Gilbey and Hill (2009) in their research, which suggested that many participants might not have been utilising the maps given for each scenario in order to make their feature selection and instead, similar to the results obtained for scenario three, participants were selecting features based on how uncommon they were. In the present research, participants illustrated a tendency towards the selection of manufactured objects such as towns, roads, and railways. It is possible that these features were subjectively judged as being more useful as a result of an individual's previous experiences and knowledge of these features and as a result, they were selected most often overall.

6.2 The Role of Experience in the Susceptibility to Confirmation Bias in Pilots

As the Ladbroke Grove train crash, Air New Zealand Flight TE901, and Comair Flight 5191 have illustrated, confirmation bias can be a major contributing factor towards accidents and incidents, resulting in significant injuries and large loss of life (Croft, 2007; Lawton & Ward, 2005; Mahon, 1984). Therefore, it is justified and necessary that confirmation bias is further investigated in specialised fields such as aviation where the outcomes of a poor decision could be catastrophic. The present research aimed to investigate the role of experience in the susceptibility to confirmation bias in pilots. As posited by Wickens and Flach (1988), a person's rate of success on a task involving complex decision-making will be greatly influenced by their ability to successfully gather the most appropriate, relevant, and salient information and ignore the irrelevant material. It is argued that an individual's ability to be able to complete this process effectively would be largely based on their experience and prior knowledge of a task (Wickens & Flach, 1988). Klayman and Brown (1993) also argued that more experienced individuals would be better at evaluating evidence and appropriately selecting information that was diagnostic as opposed to simply familiar.

Following on from previous research by McKenzie (2006) into the likelihood of increased success when individuals are placed in situations where the features and hypotheses to be tested and evaluated are more familiar to the participants, it was expected that the participants in the novice pilot group and the experienced pilot group would perform better than the control group who had no previous experience with aviation visual navigation charts or aviation lost scenarios. Therefore, it was hypothesised that experienced pilots would be less susceptible to confirmation bias than novice pilots, who in turn would be less susceptible to confirmation bias than non-pilots. It was expected that the more experienced individuals would have a greater range of schemas available to them to draw on during the hypothesis testing process, providing them with a greater range of solutions to consider.

During the hypothesis-testing process, situational cues and task features are compared to peoples' existing schemas to see if they can locate a matching experience, solution, and outcome. If a perfect match is found then the situation can be resolved and the outcome will be successful. The greater the experience of an individual in a particular field, the greater the number of schemas that they will have available to them to draw from in the search for a solution. Unfortunately, it is often difficult to find a 100% perfect match and the situational cues given may not be effective diagnostic tools, e.g., "to the lost pilot, the 60^{o} intersection of a freeway with a road below may be consistent with several different ground locations" (Wickens & Flach, 1988, p. 131). In the present research, two out of the three task features given for each scenario were consistent with both the given location hypothesis and also an alternative location which represented the person's true location. Despite prior research by Adams (1993), Madhavan and Lacson (2006), and O'Hare (2008) suggesting that experienced pilots have enhanced decision-making abilities and are less susceptible to cognitive biases than less experienced pilots, the present research found no relationship between an individual's total flight experience and their susceptibility to confirmation bias. Consequently, the hypothesis that increased levels of flight experience would reduce a person's susceptibility to confirmation bias was not supported.

As suggested by Marsh and Hanlon (2007), it is possible that because of existing knowledge and ingrained decision making patterns, individuals that are more experienced may in fact be *more* susceptible to confirmation bias. In familiar situations, an expert can draw from their existing knowledge and experience to solve complex situations. Their vast array of knowledge allows them to react quickly in scenarios involving familiar information and situational cues. In novel situations existing knowledge can become useless, and even detrimental, as an expert becomes susceptible to the same error-prone problem-solving mechanisms as novices, which are all aimed at narrowing the available solution possibilities so that the information coming in is more manageable (Beaty, 1995; Chassy & Gobet, 2011). The search space can quickly be reduced by looking for information to confirm the current hypothesis and by placing little weight or importance on contradictory information. In novel situations, individuals will generally not have a schema to draw on and the question becomes one of whether or not an individual is able to transfer their learned knowledge to solve a problem similar in nature, yet containing novel features that require a different method of action (Chen & Mo, 2004; Klayman, 1995).

Research by Bilalic et al. (2008a, 2008b, 2010) investigating the performance of experienced chess players found that high levels of experience actually served to reduce the performance of their expert players when participants were faced with a familiar chess move and were asked to evaluate alternative solutions. The outcomes of Bilalic et al's. (2008a) research suggested that susceptibility to the *Einstellung* or mental set effect was directly related to the inflexibility of an expert's knowledge schemas that had been constructed and reinforced over time. The only exception was the highest rated chess players, who appeared to not be susceptible to the mental set effect, indicating that after a certain skill level, expertise was actually beneficial. In contrast to Bilalic et al's. (2008a) research, in the present research, the effects of extremely high levels of experience did not serve to increase the performance of the experienced pilot group, and the experienced pilot group did not perform significantly worse than either the novice pilot or non-pilot groups. The current research also illustrated that the occurrence of confirmation bias was stable across all participants regardless of their group membership, gender, or age.

As suggested by Reason (1990), expertise and high levels of skill can actually make some individuals more prone to certain types of error and bias: "Novices typically make mistakes due to faulty, inadequate or missing knowledge, whereas experts are prone to slips and lapses due to the way in which skilled routines and subroutines are compiled and executed" (Reason, 1990, p. 370). In the present research, it was found that both the non-pilots and novice pilots applied a hypothesis-confirming strategy most regularly in their feature selection, and illustrated a tendency to select large features and manufactured objects. In contrast, similar to Reason's (1990) supposition, the experienced pilot group favoured both manufactured objects and large features most regularly, but not hypothesis-confirming information. Although the experienced pilot group performed just as poorly as the other two participant groups on the location discovery task, the thematic analysis identified that they were not scoring incorrectly for the same reasons. As suggested by Raab and Johnson (2007) the results indicate that knowledge of expertise levels and the exact role they play in the generation of hypotheses, assessment and evaluation of these hypotheses and how an option is chosen from among the choices produced, is still relatively unclear.

6.3 Practical Implications

The present research serves to further the exploration of confirmation bias in the field of aviation. Research in this area is relatively new and, despite recent reports by Bazargan and Guzhva (2011) and Wiegmann and Shappell (2003) demonstrating that errors in aviation decision-making are one of the largest contributors to aviation accidents and incidents, little research has systematically identified the factors contributing to the occurrence of confirmation bias in pilots. In the field of aviation it is the role of the professional pilot to efficiently and accurately collate the available facts, assess and evaluate the evidence, form a hypothesis of the issue, and make a choice as to the plan of action (Rudolph, Morrison, & Carroll, 2009). Often the information available is incomplete, the individuals may not have time to gather further information, and the decision is made under pressure. This can result in a less than ideal situation and a potentially catastrophic outcome. Pilots must be able to react quickly and accurately and remain flexible so that if a situation changes or new information becomes known, they are able to change their decision and specify a new action plan (Beaty, 1995).

Despite years of experience and training in aircraft systems, human factors, and navigation, the current research indicates that even the most experienced pilots can be susceptible to confirmation bias. As the present research illustrates, it appears that the current aviation lost procedures utilised in New Zealand aviation may in fact promote the incidence of confirmation bias and a review of these procedures could certainly be useful in reducing its occurrence. The qualifications of drawing a circle of uncertainty based only on a few features are inherently dangerous. The processes involved in this procedure should ensure that individuals attempt to find an identifying feature unique to the location they are in. The process may benefit from encouraging pilots, as part of their standard operating procedure, to critically re-evaluate all decisions made and spend some time, if practical, seeking any evidence that may disprove their current hypothesis. Banbury, Dudfield, Hoermann, and Soll (2007) also argue that all pilots should seek to confirm their current course of action by first seeking out evidence to disprove their current theory. More research will need to be completed in the usefulness and viability of implementing this process as it may not always be appropriate. Pilots may feel pressured to find a disconfirming feature where one does not exist or if the aviation visual navigation charts are out of date, not all features visible on the ground will be available on the chart to provide a cross-reference.

Under normal circumstances, in benign everyday situations, a hypothesis testing strategy that employs the use of confirming information more often than disconfirming information is actually quite reasonable and appropriate, particularly as the consequences of a wrong selection are not as costly (Friedrich, 1993) and often, positive tests of the hypothesis and target set enable disconfirmation of the hypothesis (Klayman & Ha, 1987). Fugelsang, Stein, Green, and Dunbar (2004) also illustrated that scientists studied in their natural surroundings were initially stubborn in accepting disconfirming information against their initial hypotheses, but repeated disconfirmation through research and evidence resulted in a change of theory over time. Unfortunately, in the field of aviation, pilots often have to make decisions under pressure and if the solution that is trialled is incorrect, it could result in injury or death. It is possible that gradually a pilot may accumulate disconfirming evidence that refutes their initial hypothesis about a situation, but unfortunately, due to time and pressure constraints a subsequent re-evaluation of the initial hypothesis in light of the new evidence obtained may not be possible. Therefore, it is important that the mechanisms underlying the susceptibility to confirmation bias in pilots are well understood and attempts can be made to eliminate or reduce its effects.

Unfortunately, prior research has had little success in eradicating the effects of confirmation bias from the hypothesis evaluation and testing process (Teasley et al., 1994, Gilbey & Hill, 2009). Crew Resource Management (CRM) was implemented as a training programme to identify the human error-related causes of aviation accidents and design training programmes based on these findings in an effort to reduce pilot error (Helmreich, Merritt, & Wilhelm, 1999). Unfortunately, even CRM training has failed to properly address cognitive biases and research has suggested that the impacts of CRM differ by culture and the positive effects of the initial training decay over time if the training is not followed up, or is not wholly supported by management and supervisors (Helmreich, Merritt, & Wilhelm, 1999; Murray, 1999). Lenné, Ashby, and Fitzharris (2008) posit that systems implemented to control for error will have limited success if they are only aimed at one aspect of the complex error system and any countermeasures put in place that operate on only one aspect of the complex system will invariably fail. As prior research has suggested, confirmation bias is fairly resistant to basic training on the subject and simply making people aware that confirmation bias exists is not sufficient to eliminate it (Gilbey & Hill, 2009).

6.4 Limitations

One of the limitations of the present study was the uneven number of participants who completed each of the scenario orders. To counteract the effects of order on subsequent task performance a balanced Latin square design was utilised. The experiment server was designed to administer the experiment through the different orderings each time a participant completed the experiment online. For example, participant one completed the scenarios in order one and once completed, participant number two would be assigned to complete the experiment in order two. Unfortunately, because individuals were accessing and completing the online experiment simultaneously, the number of participants completing the different orders was not equal, with a much higher proportion of participants having completed the experiment under the first order. Order one $n = 20$ (37.7%), order two $n = 11$ (20.8%), order three $n = 10$ (18.9%) and order four $n = 12$ (22.6%). To confirm that order effects were not playing a role in the results of the research, an ANOVA was completed to check for any statistical differences. No effect of order was detected and all subsequent analyses were able to ignore this factor. However, future research would ideally have an equal number of participants complete the scenarios in each of the order sequences.

The present research was also limited by sample size. Despite having a total sample size of 53 overall, which was suitable for the statistical analyses, once the participants had been split into their groups, the group sizes were significantly smaller, and the novice pilot group only had a group size of 13. Future research looking into confirmation bias in the field of aviation would benefit from a much larger sample size to allow for a greater comparison of the factors contributing to confirmation bias in pilots. Also, the task utilised in the present research was much easier than many situations that a pilot could expect to experience during the course of their piloting career and therefore the results obtained may not be a realistic representation of pilots' true performance. A further limitation is that the thematic analysis was only carried out by a single researcher and consequently the results from the thematic analysis must be interpreted with caution.

6.5 Future Research

Levels of susceptibility to confirmation bias were stable across the three participant groups indicating that future research needs to look at the actual mechanisms underlying the occurrence of confirmation bias. A lot of research has been completed in the field of cataloguing and describing bias in human judgement and decision-making, yet little has actually been completed in the way of valid models of how to reduce and deal with these biases (Lilienfeld, Ammirati, & Landfield, 2009). Nickerson (1998) also illustrated that plenty of research has been completed confirming the incidence of confirmation bias and the tendency for people to seek out hypothesis-confirming information, but little research has focused on the cause. Gilbey and Hill's (2009) research included a group of orienteers who all performed on the RDT task significantly better than both pilots and members of the general population. Further research would serve to benefit from the identification of further groups who are less susceptible to confirmation bias and a closer inspection of what makes these individuals unique.

Future research would also benefit from the utilisation of a flight simulator, which would be able to simulate a more realistic scenario and prompt a more thorough search of the surrounding terrain. Although the above research looks at pilots with aviation based scenarios, utilising aviation visual navigation charts, the experiment could be completed by individuals in the privacy and calmness of their own home. Certainly, the stress and activity levels were much less than if the individuals had to carry out the location discovery task as well as maintain safe flight under stress.

6.6 Conclusion

As Rodgers, Covington, and Jensen (1999) suggest, there will always be some individuals who are susceptible to error where others are not. It has been readily identified and illustrated over the decades that confirmation bias and other cognitive biases are pervasive and extremely robust in the face of attempts to eradicate them (Nickerson, 1998). Pilots and airline management need to be aware of the risk that biases like confirmation bias can pose and the potential risk for disaster that could occur if the bias is ignored. Unfortunately, many attempts to reduce or eliminate the effects of confirmation bias have been unsuccessful as the true nature of the mechanisms underlying the cause and triggers of confirmation bias are yet to be fully understood. Questions remain to be answered, including, why are some people highly susceptible to confirmation bias and others are not? In addition, why do the effects of years of dedicated training and experience in a specialised field not protect against a person's susceptibility to confirmation bias?

The present research illustrated that a participant's susceptibility to confirmation bias was not reduced as a result of the years of training and practice that pilots obtain in navigation, map reading, and aircraft handling. It was clear in the present research that the majority of participants who selected the incongruent feature on the location discovery task, did so as a result of actively seeking out disconfirming information. A closer examination of these individuals in order to fully understand why they answered the way they did would be beneficial to the field of aviation safety and the literature on the occurrence of confirmation bias in hypothesis testing and evaluation. Once the underlying causes of confirmation bias have been fully identified, a model of confirmation bias and how its effects can be reduced or eliminated can be constructed. This model can then be utilised in the field of aviation to advise on the most efficient and effective procedure to follow in scenarios requiring quick and accurate hypothesis evaluation and action. Identifying the underlying causes of confirmation bias in pilots will also aid in the construction of interventions that could potentially reduce or eradicate tragic occurrences such as the Erebus Disaster and Comair Flight 5191.

References

Adams, R. J. (1993). *How expert pilots think: Cognitive processes in expert decision making.* Washington, D. C.: Federal Aviation Administration.

Austerweil, J. L., & Griffiths, T. L. (2011). Seeking confirmation is rational for deterministic hypotheses. *Cognitive Science, 35*, 499-526. doi: 10.1111/j.1551-6709.2010.01161.x

Banbury, S., Dudfield, H., Hoermann, H-J., & Soll, H. (2007). FASA: Development and validation of a novel measure to assess the effectiveness of commercial airline pilot situation awareness training. *The International Journal of Aviation Psychology, 17*, 131-152. doi: 10.1080/10508410701328557

Bazargan, M., & Guzhva, V. S. (2011). Impact of gender, age and experience of pilots on general aviation accidents. *Accident Analysis and Prevention, 43*, 962-970. doi: 10.1016/j.aap.2010.11.023

Beattie, J., & Baron, J. (1988). Confirmation and matching biases in hypothesis testing. *The Quarterly Journal of Experimental Psychology, 40A*, 269-297. doi: 10.1080/02724988843000122

Beaty, D. (1995). *The naked pilot: The human factor in aircraft accidents.* Shrewsbury, England: Airlife Publishing Ltd.

Bilalic, M., McLeod, P., & Gobet, F. (2008a). Inflexibility of experts-reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology, 56*, 73-102. doi: 10.1016/j.cogpsych.2007.02.001

Bilalic, M., McLeod, P., & Gobet, F. (2008b). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition, 108*, 652-661. doi: 10.1016/j.cognition.2008.05.005

Bilalic, M., McLeod, P., & Gobet, F. (2010). The mechanism of the Einstellung (set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science, 19*, 111-115. doi: 10.1177/096372140363571

Blomberg, O. (2011). Conceptions of cognition for cognitive engineering. *The International Journal of Aviation Psychology, 21*, 85-104. doi: 10.1080/10508414.2011.537561

Bruner, J. S., & Postman, L. (1949). On the perception of incongruity: A paradigm. *Journal of Personality, 18*, 206-223. doi: 10.1111/j.1467-6494.1949.tb01241.x

Chassy, P., & Gobet, F. (2011). A hypothesis about the biological basis of expert intuition. *Review of General Psychology, 15*, 198-212. doi: 10.1037/a0023958

Chen, Z., & Mo, L. (2004). Schema induction in problem solving: A multidimensional analysis. *Journal of Experimental Psychology, 30*, 583-600. doi: 10.1037/0278-7393.30.3.583

Civil Aviation Act. (1990). Retrieved from http://www.legislation.govt.nz/act/public/1990/0098/latest/DLM214687.html

Civil Aviation Authority. (2011). *Pilot licenses and ratings – Commercial pilot license* (Advisory Circular, AC61-5). Wellington, New Zealand.

Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Journal of Human Factors, 50*, 745-754. doi: 10.1518/001872008X354183

Croft, J. (2007). Confirmation bias: Subtle but lethal. Retrieved from http://www.flightglobal.com/blogs/flight-international/2007/07/confirmation-bias-subtle-but-lethal.html

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin, 28*, 1379-1387. doi: 10.1177/014616702236869

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43*, 111-121. doi: 10.1016/0001-6918(79)90017-9

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725-747. doi: 10.1037/0003-066X.49.8.725

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305. doi: 10.1146/annurev.psych.47.1.273

Evans, J. ST. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*, 378-395. doi: 10.3758/BF03193858

Farris, H. H., & Revlin, R. (1989a). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition, 17*, 221-232. doi: 10.3758/BF03197071

Farris, H., & Revlin, R. (1989b). The discovery process: A counterfactual strategy. *Social Studies of Science, 19*, 497-513. doi: 10.1177/030631289019003005

Federal Aviation Administration. (2011). *Estimated active airmen certificates held.* Retrieved from http://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2011/

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*, 239-260. doi: 10.1037/0033-295X.90.3.239

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science, 14*, 195-200. doi: 10.1111/1467-9280.02431

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review, 100*, 298-319. doi: 10.1037/0033-295X.100.2.298

Fugelsang, J. A., Stein, C. B., Green, A. E., & Dunbar, K. N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology, 58*, 86-95. doi: 10.1037/h0085799

Gale, M., & Ball, L. J. (2006). Dual-goal facilitation in Wason's 2-4-6 task: What mediates successful rule discovery? *The Quarterly Journal of Experimental Psychology, 59*, 873-885. doi: 10.1080/02724980543000051

Gale, M., & Ball, L. J. (2009). Exploring the determinants of DG facilitation in a rule discovery task. *Thinking & Reasoning, 15*, 294-315. doi: 10.1080/13546780903040666

Gibb, R. W., & Olson, W. (2008). Classification of air force aviation accidents: Mishap trends and prevention. *The International Journal of Aviation Psychology, 18*, 305-325. doi: 10.1080/10508410802346913

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555-588. doi: 10.1037/a0015701

Hecht, H. & Proffitt, D. R. (1995). The price of expertise: Effects of experience on the water-level task. *Psychological Science, 6*, 90-95. doi: 10.1111/j.1467-9280.1995.tb00312.x

Helmreich, R. L., Merritt, A. C., Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology, 9*, 19-32. doi: 10.1207/s15327108ijap0901_2

Gilbey, A., & Hill, S. (2009, July). *Confirmation bias in general aviation lost procedures.* Paper presented at the 2009 Aviation Education and Research Conference, Blenheim, New Zealand.

Hooey, B. L., & Foyle, D. C. (2006). Pilot navigation errors on the airport surface: Identifying contributing factors and mitigating solutions. *The International Journal of Aviation Psychology, 16*, 51-76. doi: 10.1207/s15327108ijap1601_3

International Civil Aviation Organisation. (1993). *Investigation of human factors in accidents and incidents.* Montreal: ICAO.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision, 50*, 59-99. doi: 10.1023/A:1005296023424

Kareev, Y., Halberstadt, N., & Shafir, D. (1993). Improving performance and increasing the use of non-positive testing in a rule-discovery task. *The Quarterly Journal of Experimental Psychology, 46A*, 729-742. doi: 10.1080/14640749308401036

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-48. doi: 10.1207/s15516709cog1201_1

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation, 32*, 385-418.

Klayman, J., & Brown, K. (1993). Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment. *Cognition, 49*, 97-122. doi: 10.1016/0010-0277(93)90037-V

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228. doi: 10.1037/0033-295X.94.2.211

Klayman, J., & Ha, Y-W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology, 15*, 596-604. doi: 10.1037/0278-7393.15.4.596

Lawton, R., & Ward, N. J. (2005). A systems analysis of the Ladbroke Grove rail crash. *Accident Analysis and Prevention, 37*, 235-244. doi: 10.1016/j.aap.2004.08.001

Lenné, M. G., Ashby, K., & Fitzharris, M. (2008). Analysis of general aviation crashes in Australia using the human factors analysis and classification system. *The International Journal of Aviation Psychology, 18*, 340-352. doi: 10.1080/10508410802346939

Lewandowsky, S., & Kirsner, K. (2000). Knowledge partitioning: Context-dependent use of expertise. *Memory & Cognition, 28*, 295-305. doi: 10.3758/BF03213807

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science, 4*, 390-398. doi: 10.1111/j.1745-6924.2009.01144.x

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109. doi: 10.1037/0022-3514.37.11.2098

Macrae, C. (2009). Human factors at sea: Common patterns of error in groundings and collisions. *Maritime Policy & Management, 36*, 21-38. doi: 10.1080/03088830802652262

Madhavan, P., & Lacson, F. C. (2006). Psychological factors affecting pilots' decisions to navigate in deteriorating weather. *North American Journal of Psychology, 8*, 47-62.

Mahon, P. (1984). *Verdict on Erebus.* Auckland: William Collins Publishers Ltd.

Marsh, D. M., & Hanlon, T. J. (2007). Seeing what we want to see: Confirmation bias in animal behavior research. *Ethology, 113*, 1089-1098. doi: 10.1111/j.1439-0310.2007.01406.x

McDonald, J., & Stenger, J. (1993). Effects of hypothesis saliency on the use of positive and diagnostic test strategies. *Organizational Behavior and Human Decision Processes, 56*, 213-232. doi: 10.1006/obhd.1993.1052

McGregor, A. (2006, March). *Accidents, failures, mistakes and leaky buildings*. Paper presented at the National IPENZ Conference, Wellington, New Zealand.

McKenzie, C. R. M. (2003). Rational models as theories – not standards – of behavior. *Trends in Cognitive Sciences, 7*, 403-406. doi: 10.1016/S1364-6613(03)00196-7

McKenzie, C. R. M. (2006). Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias. *Memory & Cognition, 34*, 577-588. doi: 10.3758/BF03193581

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57-111. doi: 10.1017/S0140525X10000968

Müller, S. M., Garcia-Retamero, R., Cokely, E., & Maldonado, A. (2011). Causal beliefs and empirical evidence. *Experimental Psychology, 58*, 324-332. doi: 10.1027/1618-3169/a000099

Murray, S. R. (1999). FACE: Fear of loss of face and the five hazardous attitudes concept. *The International Journal of Aviation Psychology, 9*, 403-411. doi: 10.1207/s15327108ijap0904_6

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology, 46A*, 759-778. doi: 10.1080/14640749308401038

Mynatt, C. R., Doherty, M. E., Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29*, 85-95. doi: 10.1080/00335557743000053

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of a confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology, 30*, 395-406. doi: 10.1080/00335557843000007

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112*, 979-999. doi: 10.1037/0033-295X.112.4.979

Nickerson, R. S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning, 2*, 1-31. doi: 10.1080/135467896394546

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220. doi: 10.1037/1089-2680.2.2.175

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education, 44*, 94-100. doi: 10.1111/j.1365-2923.2009.03507.x

National Transportation Safety Board. (2007). *Attempted takeoff from wrong runway, Comair flight 5191* (Report NTSB/AAR-07/05). Washington, DC: NTSB.

Oaksford, M., & Chater, N. (1994). Another look at eliminative and enumerative behaviour in a conceptual task. *European Journal of Cognitive Psychology, 6*, 149-169. doi: 10.1080/09541449408520141

O'Hare, D. (2006). Cognitive functions and performance shaping factors in aviation accidents and incidents. *The International Journal of Aviation Psychology, 16*, 145-156. doi: 10.1207/s15327108ijap1602_2

O'Hare, D. (2008). Book review [Review of the book *The limits of expertise: Rethinking pilot error and the causes of airline accidents*, by R. K. Dismukes, B. Berman & L. Loukopoulos]. *International Journal of Aviation Psychology, 18*, 369-371. doi:10.1080/10508410802347085

O'Hare, D., Mullen, N., Wiggins, M., Molesworth, B. (2008). Finding the right case: The role of predictive features in memory for aviation accidents. *Applied Cognitive Psychology, 22*, 1163-1180. doi: 10.1002/acp.1428

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79-97). New York, NY: Psychology Press.

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Australia, Allen and Unwin.

Penner, D. E., & Klahr, D. (1996). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory & Cognition, 24*, 655-668. doi: 10.3758/BF03201090

Poletiek, F. H. (1996). Paradoxes of falsification. *The Quarterly Journal of Experimental Psychology, 49A*, 447-462. doi: 10.1080/027249896392720

Raab, M., Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology, 13*, 158-170. doi: 10.1037/1076-898X.13.3.158

Rassin, E. (2008). Individual differences in the susceptibility to confirmation bias. *Netherlands Journal of Psychology, 64*, 87-93. doi: 10.1007/BF03076410

Rassin, E. (2010). Blindness to alternative scenarios in evidence evaluation. *Journal of Investigative Psychology and Offender Profiling, 7*, 153-163. doi: 10.1002/jip.116

Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: An analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling, 7*, 231-246. doi: 10.1002/jip.126

Reason, J. T. (1990). *Human error.* Cambridge, UK: Cambridge University Press.

Reich, S. S., & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology, 73*, 395-405. doi: 10.1111/j.2044-8295.1982.tb01822.x

Rodgers, D. A., Covington, E. C., & Jensen, R. S. (1999). Response predispositions and piloting safety. *The International Journal of Aviation Psychology, 9*, 73-90. doi: 10.1207/s15327108ijap0901_5

Rossi, S., Caverni, J. P., & Girotto, V. (2001). Hypothesis testing in a rule discovery problem: When a focused procedure is effective. *The Quarterly Journal of Experimental Psychology, 54A*, 263-267. doi: 10.1080/02724980042000101

Rudolph, J. W., Morrison, J. B., & Carroll, J. S. (2009). The dynamics of action-oriented problem solving: Linking interpretation and choice. *Academy of Management Review, 34*, 733-756. doi: 10.5465/AMR.2009.44886170

Schwind, C., Buder, J., Cress, U., & Hesse, F. W. (2012). Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating thinking. *Computers & Education, 58*, 787-796. doi: 10.1016/j.compedu.2011.10.003

Shappell, S. A., & Wiegmann, D. A. (1997). A human error approach to accident investigation: The taxonomy of unsafe operations. *The International Journal of Aviation Psychology, 7*, 269-291. doi: 10.1207/s15327108ijap0704_2

Silverman, B. G. (1992). Modeling and critiquing the confirmation bias in human reasoning. *IEEE Transactions on Systems, Man, and Cybernetics, 22*, 972-982. doi: 10.1109/21.179837

Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482-488. doi: 10.1126/science.183.4124.482

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Psychology, 22*, 93-121. doi: 10.1016/0022-1031(86)90031-4

Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory & Cognition, 20*, 392-405. doi: 10.3758/BF03210923

Spellman, B. A., López, A., & Smith, E. E. (1999). Hypothesis testing: Strategy selection for generalising versus limiting hypotheses. *Thinking and reasoning, 5*, 67-91. doi: 10.1080/135467899394084

Teasley, B. E., Leventhal, L. M., Mynatt, C. R., & Rohlman, D. S. (1994). Why software testing is sometimes ineffective: Two applied studies of positive testing strategy. *Journal of Applied Psychology, 79*, 142-155. doi: 10.1037/0021-9010.79.1.142

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research, 40*, 271-300. doi: 10.1177/1046496409332928

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1-10. doi: 10.2307/1129583

Tukey, D. D. (1986). A philosophical analysis of subjects’ modes of inquiry in Wason’s 2-4-6 task. *The Quarterly Journal of Experimental Psychology, 38A*, 5-33. doi: 10.1080/14640748608401583

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32*, 109-123. doi: 10.1080/00335558008248237

Vallée-Tourangeau, F., Austin, N. G., & Rankin, S. (1995). Inducing a rule in Wason’s 2-4-6 task: A test of the information-quantity and goal-complementarity hypotheses. *The Quarterly Journal of Experimental Psychology, 48A*, 895-914. doi: 10.1080/14640749508401422

Vallée-Tourangeau, F., & Payton, T. (2008). Graphical representation fosters discovery in the 2-4-6 task. *The Quarterly Journal of Experimental Psychology, 61*, 625-640. doi: 10.1080/17470210701268039

Vallée-Tourangeau, F., & Penney, A. K. (2005). The impact of external representation in a rule discovery task. *European Journal of Cognitive Psychology, 17*, 820-834. doi: 10.1080/09541440440440000249

Van der Henst, J., Rossi, S., & Schroyens, W. (2002). When participants are not misled they are not so bad after all: A pragmatic analysis of a rule discovery task. In W. D. Gray., & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, (pp. 902-907). Mahwah, New Jersey: Lawrence Erlbaum.

Van Swol, L. M. (2007). Perceived importance of information: The effects of mentioning information, shared information bias, ownership bias, reiteration, and confirmation bias. *Group Processes & Intergroup Relations, 10*, 239-256. doi: 10.1177/1368430207074730

Vette, G. (1983). *Impact Erebus.* Hong Kong: Hodder and Stoughton Ltd.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology, 12*, 129-140. doi: 10.1080/17470216008416717

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology, 20*, 273-281. doi: 10.1080/14640746808400161

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*, 63-71. doi: 10.1080/00335557143000068

Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *The Quarterly Journal of Experimental Psychology, 14*, 246-249. doi: 10.1080/17470216208416542

Wharton, C. M., Cheng, P. W., Wickens, T. D. (1993). Hypothesis testing strategies: Why two goals are better than one. *The Quarterly Journal of Experimental Psychology, 46A*, 743-758. doi: 10.1080/14640749308401037

Wickens, C. D., & Flach, J. M. (1988). Information processing. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 111-155). San Diego, CA: Academic Press.

Wiegmann, D. A., & Shappell, S. A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology, 7*, 67-81. doi: 10.1207/s15327108ijap0701_4

Wiegmann, D. A., & Shappell, S. A. (2001). Human error perspectives in aviation. *The International Journal of Aviation Psychology, 11*, 341-357. doi: 10.1207/S15327108IJAP1104_2

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system.* Hants, England: Ashgate Publishing Ltd.

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*, 716-730. doi: 10.3758/BF03211392

Appendix A

Appendix A. Information Sheet.

The Role of Experience in the Susceptibility to Confirmation Bias in Pilots

INFORMATION SHEET

My name is Jaime Rowntree and I am a Master of Arts student in the School of Psychology at Massey University. I am conducting a research experiment, which aims to identify differences in patterns of decision-making between non-pilots, novice pilots, and experienced pilots and also investigate the role of experience in the susceptibility to confirmation bias in pilots.

Please take the time to read this information sheet carefully and consider whether you would like to participate in this research. Please contact me using the details listed below if you have any questions or if you would like some more information. Thank you for reading this information sheet.

Who can take part in this research?

I am interested in receiving responses from a wide range of individuals. This includes student and trainee pilots, recreational pilots, and individuals who hold a pilot's license of any level. This research project is open to all pilots of aircraft in the following categories, un-powered fixed wing (Glider), powered fixed-wing (Aeroplane), and powered rotary-wing (Helicopter) pilots. I am also interested in the differences that exist in the patterns of decision making between pilots and non-pilots and would also like to receive responses from participants who have no previous flying experience.

What does the experiment involve?

This research experiment consists of some initial questions to ascertain your experience in flying an aircraft. This is then followed by four aviation navigation scenarios that will ask you to imagine yourself in four situations where you are not certain of your location. For each situation, you will be asked to make a judgement about your location on a map given some information about what identifiable features you can see. The time to complete the research experiment will be approximately 5 to 10 minutes.

Your responses are anonymous. Completion of the online research experiment implies consent. You have the right to decline to answer any particular question.

If you are happy to continue, you can **click the link in the box below** to take you through to the online research experiment. If you have any further queries or would like to know a little bit more about the study before you participate, please feel free to contact one or all of the following people at the bottom of this page.

Your Rights

Completion and submission of the following research experiment implies your consent to participating in the research.
You have the right to decline to answer any particular question.

Please **CLICK HERE** if you would like to continue and participate in this research.

Project Contacts

Please feel free to contact one or all of the following people if you have any questions in regards to the research being undertaken.

Jaime Rowntree:
Telephone: 09 550 0587
Email: jaime.rowntree.1@uni.massey.ac.nz

Dr. Stephen Hill, Research Supervisor:
School of Psychology
Massey University, Private Bag 11 222, Palmerston North 4442,
Telephone: 06 356 9099 ext. 7566
Email: s.r.hill@massey.ac.nz

"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named above are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher(s), please contact Professor John O'Neill, Director, (Research Ethics), telephone 06 350 5249, email humanethics@massey.ac.nz".

Appendix B

Appendix B. Email Invitation to Participants.

Hi,

My name is Jaime Rowntree and I am a current Master of Arts (Psychology) student at Massey University. I am currently recruiting participants for my research into the role of experience in the susceptibility to confirmation bias in pilots. I am also interested in how patterns of decision making differ between pilots and those with no previous flying experience.

Individuals with any number of logged flying hours can participate. This includes student, trainee, private, recreational, commercial and professional pilots. As I am also interested in the patterns of decision making between pilots and non-pilots, I am also interested in receiving responses from individuals with no previous flying experience.

The research can be accessed and completed online and will take five to ten minutes to complete. If you are interested in participating, please follow the link below. This link will direct you to the information sheet for the research. To ensure that the experiment displays correctly it is best to have your browser window maximised to full screen. You will also need to have JavaScript enabled (just about everybody already has this). Please also avoid using the back button on your browser, as it will take you back to the beginning of the experiment.

Thank you in advance for assisting me in the completion of my research through your participation. It is much appreciated. Please feel free to pass this research invitation on. If at any time you have any questions you may contact me or my supervisor Dr. Stephen Hill. Our contact details are listed in full on the information sheet.

Please follow the link below:

http://psychlab.massey.ac.nz:8080/psych/ajax/ConfirmationBias2.jsp

Once again thank you for taking the time to read this research invitation.

Kind Regards

Jaime Rowntree

Appendix C

Appendix C. Scenarios Used in the Location Discovery Task.

Scenario One Description

Imagine that you are the pilot in command of a light aircraft on route from Napier to Auckland. You are currently flying at 2000 feet at a speed of 110kts. You checked the weather report before taking off from Napier at 10:30am this morning and everything looked fantastic for a great flight.

Five minutes after crossing Lake Taupo you encounter some heavy thunderstorm clouds to your North. Visibility is rapidly reducing and you make the decision to try and fly around the weather. You reduce your altitude to 1500ft to stay below the clouds. Ten minutes later you have managed to fly around the clouds and you now have clear visibility to your North. Unfortunately during your fly around you have deviated off course and are no longer 100% sure of your location.

If you had not encountered bad weather you would roughly know where you should be now and you draw a circle over the area (I have completed this task for you, but for the purposes of this scenario please imagine that you drew it yourself). It is vital that you work out as soon as possible whether or not you are actually situated in the circle you have drawn to ensure that you can divert to the nearest aerodrome if the weather closes in again.

If you are in fact located in the area circled on your map, then what you can see on the ground from your aircraft should agree with the features in and around the circle drawn on your map. For example, if you can see a town then you would expect a town to be in or close to the circle drawn on your map. You can also see things beyond the area circled on the map but not to the South as a result of the weather.

Listed below are three things that you can currently see from your aircraft. Choose the one that you feel is most useful in deciding whether you are in the area circled on the map by clicking on one of the buttons on the next screen.

1. **There is a town directly below you (shown as an area of bright yellow shading).**

2. **The road (shown as a brown line) and the railway (shown as a black line) run side by side below you.**

3. **There is a large pine forestry area to the east, west and north of you.**

When you feel comfortable with your understanding of the scenario above, you can click the button at the bottom of the screen to see the map and make your decision. The descriptions of the three features above will be shown again at the right of the map.

Scenario Two Description

Imagine that you are the pilot in command of a light aircraft on route from Palmerston North to New Plymouth. You have one passenger on board who is a close friend of yours. You are currently flying at 1500 feet at a speed of 105kts.

About fifteen minutes from New Plymouth your friend starts to feel ill and begins complaining of chest pains. Their condition quickly deteriorates and you try to help your friend as best you can while still maintaining control of your aircraft. You realise that you need to land as soon as possible and get your friend medical help. Unfortunately while you have been trying to assist your friend you have deviated off course and are no longer 100% sure of your location.

If your friend had not fallen ill you would roughly know where you should be now and you draw a circle over the area (I have completed this task for you, but for the purposes of this scenario please imagine that you drew it yourself). It is vital that you work out as soon as possible whether or not you are actually situated in the circle you have drawn to ensure that you can divert to the nearest aerodrome to seek medical attention for your friend.

If you are in fact located in the area circled on your map, then what you can see on the ground from your aircraft should agree with the features in and around the circle drawn on your map. For example, if you can see a town then you would expect a town to be in or close to the circle drawn on your map. You can also see things beyond the area circled on the map.

Listed below are three things that you can currently see from your aircraft. Choose the one that you feel is most useful in deciding whether you are in the area circled on the map by clicking on one of the buttons on the next screen.

1. There is a cluster of buildings which looks like a small town to your West (shown as an area of bright yellow shading).

2. The road (shown as a brown line) and the railway (shown as a black line) run side by side below you in an east-west direction.

3. There is a tall mountain summit to your West.

When you feel comfortable with your understanding of the scenario above, you can click the button at the bottom of the screen to see the map and make your decision. The descriptions of the three features above will be shown again at the right of the map.

Scenario Three Description

Imagine that you are the pilot in command of a light aircraft on route from Taupo to Whitianga. You are currently flying at 1500 feet at a speed of 115kts. Weather is perfect for flying and visibility conditions are superb.

As you approach Matamata you switch over to their radio frequency and are informed that they currently have a glider competition occurring today. You make the decision to divert around Matamata to try and avoid the gliders. Unfortunately you are not familiar with this area and as you make your diversion off your planned route you realise that you are no longer 100% sure of your location.

If you had not diverted around Matamata you would roughly know where you should be now and you draw a circle over the area (I have completed this task for you, but for the purposes of this scenario please imagine that you drew it yourself). It is vital that you work out as soon as possible whether or not you are actually situated in the circle you have drawn to ensure that you continue to avoid the gliders and also get back on track for your destination.

If you are in fact located in the area circled on your map, then what you can see on the ground from your aircraft should agree with the features in and around the circle drawn on your map. For example, if you can see a town then you would expect a town to be in or close to the circle drawn on your map. You can also see things beyond the area circled on the map.

Listed below are three things that you can currently see from your aircraft. Choose the one that you feel is most useful in deciding whether you are in the area circled on the map by clicking on one of the buttons on the next screen.

1. **There is an area of mountainous ranges to your East.**
2. **There is an area of swamp and marshlands directly below you.**
3. **There is a town to your East (shown as an area of bright yellow shading).**

When you feel comfortable with your understanding of the scenario above, you can click the button at the bottom of the screen to see the map and make your decision. The descriptions of the three features above will be shown again at the right of the map.

Scenario Four Description

Imagine that you are the pilot in command of a light aircraft on route from Invercargill to Dunedin. You are currently flying at 1500 feet at a speed of 105kts. There are a few clouds about which are causing some mild turbulence, visibility is fair.

After flying in a North Easterly direction for twenty five minutes you pass through the town of Gore and decide to change your heading to East so that you can fly out to the coast and then North up to Dunedin to get a great scenic view of the rugged New Zealand coastline. Fifteen minutes out of Gore your engine splutters and your engine loses power. You notice that your oil pressure gauge is also going down and you immediately try and work out what may be causing the engine problems. Unfortunately during this time you lose track of where you are and you realise that you are no longer 100% sure of your location.

If you had not been distracted by your aircraft's engine trouble you would roughly know where you should be now and you draw a circle over the area (I have completed this task for you, but for the purposes of this scenario please imagine that you drew it yourself). It is vital that you work out as soon as possible whether or not you are actually situated in the circle you have drawn to ensure that you can divert to the nearest aerodrome to make an emergency landing.

If you are in fact located in the area circled on your map, then what you can see on the ground from your aircraft should agree with the features in and around the circle drawn on your map. For example, if you can see a town then you would expect a town to be in or close to the circle drawn on your map. You can also see things beyond the area circled on the map.

Listed below are three things that you can currently see from your aircraft. Choose the one that you feel is most useful in deciding whether you are in the area circled on the map by clicking on one of the buttons on the next screen.

1. **The road (shown as a brown line) and the railway (shown as a black line) run side by side below you.**

2. **There are mountainous ranges to your West.**

3. **There is a large area of pine forest to your East.**

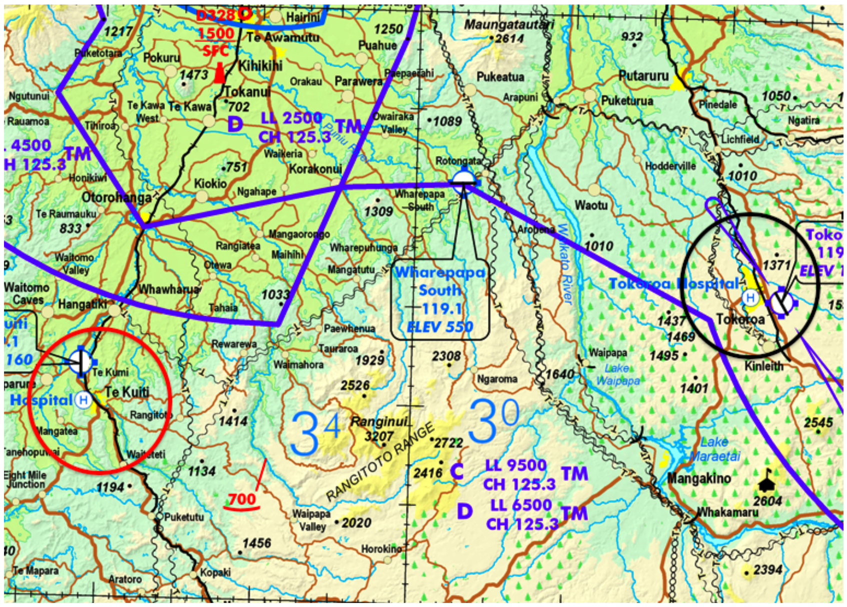
When you feel comfortable with your understanding of the scenario above, you can click the button at the bottom of the screen to see the map and make your decision. The descriptions of the three features above will be shown again at the right of the map.

Appendix D

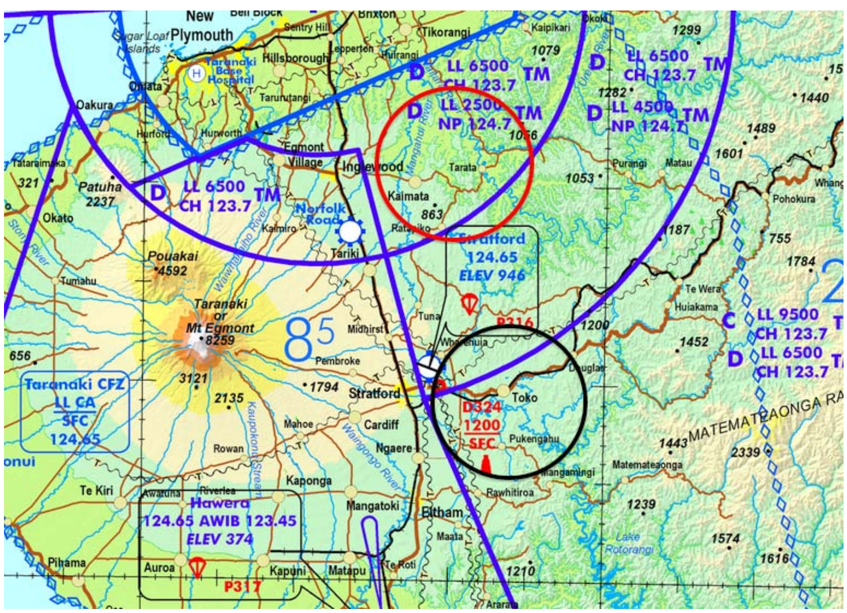
Appendix D. Aviation Visual Navigation Charts.

The red circle represents R_H The black circle represents R_T

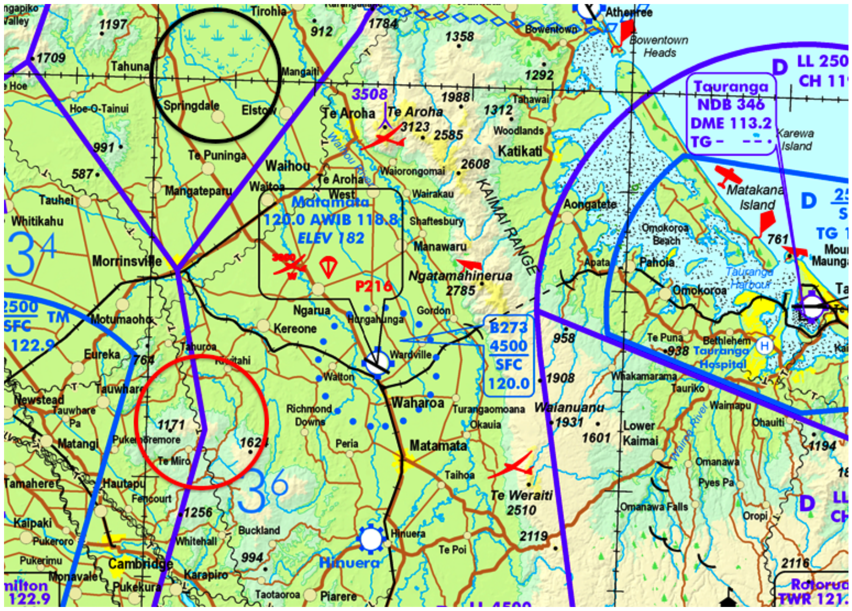
Scenario 1: Aviation Visual Navigation Chart – Master Copy



Scenario 2: Aviation Visual Navigation Chart – Master Copy



Scenario 3: Aviation Visual Navigation Chart – Master Copy



Scenario 4: Aviation Visual Navigation Chart – Master Copy