The Deliberation Time-Sink:

The Rationality of Gathering Information

by

Laurel Evans

Thesis submitted to the School of Psychology, Cardiff University,

for the degree of

DOCTOR OF PHILOSOPHY

September 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

First and foremost, I would like to thank my family. Without their enduring support and encouragement, I could never have undertaken a Ph.D.

The single individual who deserves the most thanks is Dr. Marc Buehner, my Ph.D. supervisor. Somehow he has managed to provide not only pointed and excellent guidance, but also an incredible freedom to find my own path (and occasionally even make mistakes). He has been far kinder to me than I deserve. I would also like to thank Dr. Merideth Gattis, who, while not my Ph.D. supervisor, provided me with an early opportunity to join in research with her and thus learn first-hand the methods of a different research team. I have often reflected on the value of such a broad base of training. Finally, I thank Ella, their daughter, for her patience in allowing me to steal away her parents' time.

There are several other people who have given me help or advice throughout my study. Special thanks go to Andy Maul and Siân Jones for their many bits of cogent advice on statistics. I'd also like to thank Dina Dosmukhambetova, Rachael Elward, Joseph Sweetman, Michael Lewis, Ulrike Hahn, Dominic Dwyer, James Close, Greg Maio, Dan Turetsky, Gruffydd Humphreys, Jacky Boivin, and Todd Bailey. Thanks also to the entire Cognitive group for listening to me present my work and offering supportive feedback. Finally, Gruffydd Humphreys, Adam Cassar, and Lauren Constable assisted me with data collection for Experiments V/VI and two pilot experiments, respectively, and deserve large thanks for this.

There were also many people who provided great friendship, listening to my problems, encouraging me, and cheering me on throughout my work – these people even occasionally forced me to take breaks. They are too numerous to list in full, but I would like especially to acknowledge Dan Turetsky, Andy Maul, Andrew Cheng, Gruffydd Humphreys, and Tim Schofield.

Finally, I'd like to acknowledge the teachers who inspired me from a young age to pursue my questions, and to pursue rationality. To all my mathematics teachers, especially Stephen Maurer. To my psychology teachers, especially Frank Durgin and Barry Schwartz. And finally, to the man who first drove the question "Why?" into my head, through repeated application: to Stan Heard. Thank you all.

Related Work

The work in this thesis has been presented at several international conferences and workshops. I presented the poster "Thinking Style and the Accuracy/Efficiency Trade-off" at the European Society for Cognitive Psychology XV Conference in Marseille, France, in August of 2007, the paper "Choice and Wisdom" at the Summer School on Decision Processes in Milan, Italy, in July of 2008, the poster "Individual Differences in Sampling Behaviour" at the International Conference on Thinking in Venice, Italy, in August of 2008, and the paper "Confidence Growth During Information Gathering" at the Human Problem Solving Workshop in West Lafayette, Indiana, USA, in November of 2008. Finally, my work on determining the rationality of taking small samples led me to publish a comment, now in press at the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, entitled "Small Samples Do Not Cause Greater Accuracy – But Clear Data May Cause Small Samples. Comment on Fiedler and Kareev (2006)."

Contents

ABSTRACT 1

1. INTRODUCTION 3

1.1. RATIONALITY AND DECISION MAKING 3

1.2. INDIVIDUAL DIFFERENCES IN RATIONALITY: THINKING STYLES AND NEED FOR COGNITION 10

1.3. DUAL AND TRI-PROCESS THEORIES: NFC'S ROLE IN COGNITION 14

1.4. TEACHING RATIONALITY 21

1.5. POTENTIAL DRAWBACKS: IS DEEP THINKING ALWAYS RATIONAL? 22

1.6. SAMPLING-BASED CHOICE: A BRIEF INTRODUCTION 24

1.7. SUMMARY AND EXPERIMENT PLAN 28

1.7.1. Analysis overview: Assumptions and analysis procedures used for all experiments *29*

2. EXPERIMENT I: A LONG SEARCH FOR INFORMATION **31**

2.1. METHOD 33

2.1.1. Participants *33*

2.1.2. Procedure *33*

2.1.3. Choice task *34*

2.1.4. Materials *36*

2.1.4.1. Cognitive ability measure: Raven's Progressive Matrices 36

2.1.4.2. Need for Cognition (NFC) scale 36

2.1.4.3. Social Desirability (SDR) scale 37

2.2. RESULTS 38

2.2.1. Descriptive statistics *38*

2.2.2. Main analyses *40*

2.3. DISCUSSION 46

2.3.1. Ability failure *47*

2.3.2. Strategy differences: Accuracy focus *48*

2.3.3. Strategy differences: Improvement *50*

2.3.4. Strategy differences: Pattern-searching *51*

2.3.5. Motivation differences: Boredom *53*

2.3.6. Motivation differences: Social desirability 53

2.3.7. A heuristic for spending time 55

2.3.8. Summary 55

3. EXPERIMENT II: BECOMING MORE EFFICIENT **57**

3.1. METHOD 58

3.1.1. Participants 58

3.1.2. Procedure and choice task 58

3.1.3. Materials 59

3.2. RESULTS 59

3.2.1. Main analyses 63

3.3. DISCUSSION 73

4. EXPERIMENT III: THRESHOLD DETERMINATION **79**

4.1. METHOD 81

4.1.1. Participants 81

4.1.2. Design 81

4.1.3. Procedure 83

4.1.4. Materials 84

4.2. RESULTS 84

4.2.1. Main analyses 86

4.3. DISCUSSION 88

5. EXPERIMENT IV: THE ACCUMULATION OF CONFIDENCE **92**

5.1. METHOD 94

5.1.1. Participants 94

5.1.2. Procedure 94

5.1.3. Choice task 94

5.1.4. Materials 96

5.2. RESULTS 96

5.2.1. Main analyses 99

5.2.1.1. Confidence 100

5.3. DISCUSSION 104

6. EXPERIMENT V: A FURTHER TEST FOR ACCURACY FOCUS 106

6.1. METHOD 108

6.1.1. Participants 108

6.1.2. Procedure 108

6.1.3. Choice task 109

6.1.4. Materials 109

6.2. RESULTS 110

6.2.1. Main analyses 115

6.3. DISCUSSION 119

7. EXPERIMENT VI: PERFORMANCE ON A SMALL-SCALE SAMPLING TASK: THE TRIVIA PROBLEM 122

7.1. METHOD 123

7.1.1. Participants 123

7.1.2. Design 124

7.1.3. Procedure 126

7.1.4. Materials 128

7.2. RESULTS 128

7.2.1. Main analyses 130

7.3. DISCUSSION 135

8. OVERVIEW AND GENERAL DISCUSSION 137

8.1. SUMMARY OF EVIDENCE 139

8.1.1. Ability failure 139

8.1.2. Accuracy focus 140

8.1.3. Improvement 143

8.1.4. Pattern-searching 144

8.1.5. Boredom 144

8.1.6. Social desirability 145

8.1.7. Time heuristic 147

8.1.8. Overall *147*

8.2. DISCUSSION 148

8.2.1. Practical implications *150*

8.2.1.1. Teaching rationality for sampling-based choice 150

8.2.1.2. Teaching meta-strategies 151

8.3. FUTURE RESEARCH 151

8.3.1. Ability failure *151*

8.3.2. Accuracy focus *152*

8.3.3. Extraneous thoughts *154*

8.3.3.1. Improvement 154

8.3.3.2. Pattern-searching 155

8.3.3.3. Unrelated thoughts/Boredom 155

8.3.4. Social desirability *156*

8.3.5. Time heuristic *157*

8.3.6. New variables *158*

8.3.7. New questions *160*

8.4. CONCLUSIONS 161

REFERENCES **163**

APPENDIXES **171**

APPENDIX A: NEED FOR COGNITION SCALE 171

APPENDIX B: INSTRUCTIONS FOR EXPERIMENTS I AND II 172

APPENDIX C: INSTRUCTIONS FOR EXPERIMENT III 175

APPENDIX D: INSTRUCTIONS FOR EXPERIMENT IV 176

APPENDIX E: INSTRUCTIONS FOR EXPERIMENT V 178

APPENDIX F: INSTRUCTIONS FOR EXPERIMENT VI 179

ABSTRACT

A key feature of rationality is the use of an optimal (or *normative*) strategy, i.e. the strategy that is most likely to maximize the fulfillment of one's goals. Numerous such strategies have been explored in the literature across a wide range of problems, and many researchers have argued that humans approach several problems irrationally. Recently, researchers have begun to study individual differences in rational responding to these tasks, crucially discovering that both fluid intelligence and individual thinking styles – such as the Need for Cognition (NFC) – contribute uniquely to rational performance. Fluid intelligence is proposed as one's capacity for manipulating information in a slow, serial manner, while NFC plays a role in the engagement of such deliberative thinking.

Although the problems studied typically benefit from this type of thinking, I explore whether there might be a problem for which deliberative thinking is non-normative: sampling-based choice, in which the participant must gather information for two options before deciding which is better. I first demonstrate in two experiments that higher NFC is related to spending more time at this task – without any significant gain in accuracy – and this relationship is separate from fluid intelligence. In subsequent experiments, I explore the potential reasons for this relationship. I search for, but find no evidence that it is due to ability failure, and clear evidence shows that it is not due to boredom. I find some evidence that those high in NFC tend to focus more on accuracy. Finally, I find that the NFC-time relationship is partially mediated by social desirability, i.e. the tendency to try to promote a positive impression of one's self. Overall, this excessive focus on accuracy, and the mediator role of social desirability, suggest NFC is related to irrational performance on this task.

1. Introduction

1.1. Rationality and decision making

All people have goals. These may be as simple and fleeting as "to have a cup of tea right now" or as complex and time-consuming as "to achieve a Ph.D." They may exist at varying levels of consciousness, and they may (rightly or wrongly) receive different levels of attention and pursuit. Importantly, if we define the successful achievement of a goal in terms of the amount of satisfaction – or, in the vocabulary of the literature, *utility* (Von Neumann & Morganstern, 1944) – it brings a person within a given context, then there is exactly one way to pursue a goal that will result in the maximum or best possible outcome. The strategy that will tend to produce this best possible outcome is called the *normative* strategy (J. Baron, 2000). Note that this strategy is not merely the happenstance sequence of events that would have led to the most utility if the user were clairvoyant. E.g., if I had known it would rain on Monday and Tuesday but not Wednesday, I could have brought an umbrella on the first two days but left it at home the third, saving me the trouble of carrying it. Since I am not clairvoyant, the best or normative strategy, to maximize the happiness I get from dryness and minimize the annoyance of carrying an umbrella, would probably involve doing a good job of judging the probability that it will rain on a given day. To adopt a normative strategy in pursuing one's goals is widely viewed, along with having appropriate goals and having beliefs based on evidence, as one of the key characteristics of rationality (West, Toplak, & Stanovich, 2008).

When and where humans do *not* adopt a normative strategy, then, has become a matter of great discussion. An enormous literature has arisen on human cognitive biases – that is, tendencies to apply cognitive rules in a certain way. In particular,

there has been a large focus on the way people apply a rule of thumb (or *heuristic*) in a non-normative way. A classic example is the conjunction fallacy, as studied by Tversky and Kahneman (1983). They famously presented participants with options regarding Linda, a fictional woman, who was described as, "31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." Options given to participants included, "Linda is a bank teller" (T), "Linda is active in the feminist movement" (F), and "Linda is a bank teller who is active in the feminist movement" (T&F). Tversky and Kahneman instructed participants to rank the options according to the probability that Linda is a member of that class. The great majority (85%) of participants produced the ordering T&F > T. Tversky and Kahneman argued that this was a fallacy, given that a conjunction (T&F) is always either equally or less likely than any one of its components alone (for, considering the case that Linda is a bank teller and not active in the feminist movement (T&¬F), we can see that T = T&F + T&¬F). They proposed that participants adopted a heuristic, dubbed the representativeness heuristic, by which they judged Linda. This heuristic involves the assessment of the degree of correspondence between an item (e.g., Linda) and a population (e.g., feminists, bank tellers); i.e., how much the item represents a typical member of the population. Because Linda was more representative of a "feminist bank teller" than a "bank teller" (who may be feminist or not), the conjunction was deemed more likely. Tversky and Kahneman enumerated a number of other such heuristics that they argued produced non-normative answers in certain situations (Tversky & Kahneman, 1971, 1973, 1974). For example, the availability heuristic is used when people predict the frequency of an event (such as rain), or the proportion of a population (such as a

minority ethnic group), by the ease of which an example comes to mind. Another example is the anchoring and adjustment rule, which says that people will anchor to any recently presented number and make an adjustment to reach an answer from there, although this adjustment is usually insufficient. When Tversky and Kahneman asked, "What is the percentage of African countries in the UN?", people who were presented with a higher random number gave higher responses than people who were presented with a lower random number, and in general people do not adjust far enough from their anchor to reach the right answer (Epley & Gilovich, 2004). In each of these cases, Tversky and Kahneman argued that the heuristic people used, in the situations studied, was non-normative – that is, in many important cases, it did not produce the correct answer. People were behaving irrationally.

Another famous and oft-cited example of (supposed) human irrationality is the Wason card selection task (Wason, 1966; Wason & Johnson-Laird, 1972); the results of human performance on this task have been used to support claims that people have difficulty thinking scientifically. In the task, the participant is presented with a series of cards, such as cards with the following letters and numbers on them: A, 4, K, 7. They are then asked to identify whether a certain rule is true, such as: "If there is a vowel on one side of the card, then there must be an even number on the other side." The goal is to do this by turning over a selection of cards, ideally a minimal selection. For the example given, most people choose to turn over the "A" card (correctly) in addition to the "4" card, but importantly only a small minority (less than 25% across numerous experiments involving this task, Cosmides & Tooby, 1992), achieve the best answer: to turn over only the "A" card and the "7" card (for if there is a vowel on the other side of the "7" card, this would disprove the rule). Wason suggested that the reason participants failed to check the "7" card was that they have a bias toward

verification or confirmation (now called the *confirmation bias*), such that they prefer to establish that the statement is "true" by selecting cards that coincide with its truth. Meanwhile, they ignore searching for evidence that will falsify the hypothesis.

Naturally, there is debate over whether the behaviour on the tasks I have outlined is truly non-normative (for a review of this large literature, see Stanovich, West, & Toplak, 2010). Stanovich et al. outline the position of two dissenting camps of researchers. The first, whom they term the Meliorists, argue as above, that humans behave irrationally when they fail at tasks like the Wason selection task. The second, whom they term the Panglossians, largely hold the view that natural human performance (*descriptive* performance) is the ideal of rationality – the reason for errors on the tasks is something other than a systematic failure of analysis. It may be a temporary failure (e.g., a lapse in memory), a computational limitation, a misinterpretation of the task, or an error on the part of the psychologist about what constitutes normative (for a good example of a debate about the normative response to a famous problem, see Hahn & Warren, 2009). Stanovich et al. provide neat counters to the first two Panglossian suggestions and generally argue strongly for the Meliorist position. Temporary failures, they suggest, would not result in the high inter-correlations seen for performance both within and between reasoning tasks in the literature. Further, they claim, there is little empirical evidence for computational limitations in most cases, as measured by cognitive ability.

I concur with the Meliorists when they say that human performance, by itself, is not enough to determine rationality. Certainly, there are cases when people later admit to dissatisfaction with their decisions or their reasoning processes. However, we must be exceedingly careful in applying our definition of rationality, especially in the realm of heuristics; the Panglossians are correct to highlight the possibilities of

misinterpretation and differences in judgments of normativity. First, formal or abstract reasoning problems are certainly often different from everyday problems. For example, performance on the Wason selection task is often facilitated when the rule is *thematic* (realistic) in addition to *deontic* (involving permission), as with, "If one is to drink alcohol, then one must be over 18" (for a review, see J. S. B. T. Evans, 2007). Second, even disregarding the usual debates over whether a strategy is normative, any heuristic may be a good solution in a majority of cases in a person's life, and only occasionally produce errors; taking the long view, then, this solution may still be normative, especially if it saves one time (which is often, after all, of key importance in satisfaction). Gigerenzer and Brighton (2009) take this view, arguing that heuristics are "fast and frugal"; they also point out that many heuristics do not produce biases – instead, these not only save time but are also more accurate than their computationally intensive alternatives. Finally, it is also important to account for the relative utility of different solutions; this will change by person. In other words, although I might be able to think about the probability of rain each day and use this to determine whether I should carry an umbrella, perhaps I would be overall more satisfied if I saved that thinking time for something else and simply always carried an umbrella. Meanwhile, my friend Dan might be more annoyed by carrying an umbrella and find it worthwhile to make this judgment each day. I suggest that a person makes a rational or *wise* decision when she maximizes the fulfillment of *all* of her goals; these may or may not be the subset defined by the experimenter.

However, regardless of whether a strategy is normative in the long term or can be deemed normative for all, it is clear that, if we make some reasonable assumptions about utility and then define a rational or normative strategy to be the strategy that produces the best possible result *for that problem alone*, then there are many problems

for which people do not use the normative strategy. Oaksford and Chater (1998b) take special note of the distinction between, as they term them, local and global goals, and so we might call a strategy that maximizes utility within a defined problem context a *local-normative* strategy. For example, with the umbrella problem, if we restrict it such that we consider utility to be removed *only* by "getting wet" or "carrying something" (and not by, say, "time spent judging"), then there is a strategy that could be applied by any person in order to maximize his or her personal utility: judge the probability of rain on a given day and balance this with the relative benefit (dryness) and cost (heaviness) of carrying an umbrella. We could also call the successful resolution of the Wason selection task, with the card choices of "A" and "7", a local-normative solution. Within the carefully defined world of Wason's problem, that is clearly the correct answer. Oaksford and Chater, like myself, argue that rationality is defined in terms of global rather than local goal optimisation, and suggest the importance of considering potential global goals that participants may be optimising (though sometimes, naturally, these may coincide with local goals). I agree, but note that there are several cases in which it is important to examine the propensity for local optimisation. If, for example, I am hiring someone to solve a problem for me, I will generally want someone who is able to employ the local-normative solution – that is, to solve my problem in the best possible way. In general, any society that wishes to behave wisely (optimising goals of, say, survival and happiness of its members) may need problems solved in a local-normative way, even if such local-normativity is not optimal for individual members.

Importantly, there is also a distinction between participant-defined and experimenter-defined normativity. If, for example, an experimenter requests that a participant make choices with an equal emphasis on accuracy and speed, the

experimenter is pre-setting the utility of each, and thus experimenter-defined normativity is determined by obtaining a good ratio of accuracy to speed. Meanwhile, the participant may value accuracy over speed, such that a lower ratio is acceptable to him if it means he has high accuracy. Most experiments assess experimenter-defined normativity, purely because the participant always behaves in a participant-normative way – he will always attempt to optimise his own (local) goals, as this is precisely the process of making a choice! For this reason, participant-defined normativity is only meaningful in a global sense, as the participant struggles to optimise global goals in the face of competing local goals. Thus, if one is examining local normativity, it only makes sense to define it in terms of experimenter-defined local normativity. The study of whether certain participants can behave according to such an experimenter-defined, locally normative strategy is the focus of this thesis. This then, is what I refer to when I use the words normativity and rationality in future sections.

One further thing is clear from the history of research into rationality: Some people employ normative strategies, and some people do not. Slovic and Tversky (1974) argued that more reflective and engaged participants are more likely to understand – and thus accept – an underlying rule that produces the normative response, and are therefore more likely to come up with the solution defined by the experimenter (which is also typically the answer provided by experts). Stanovich and West (1999) dubbed this argument the understanding/acceptance assumption, and they and others have proceeded in earnest to gather data on individual differences in rational thought.

1.2. Individual differences in rationality: Thinking styles and Need for Cognition

A thinking style, also commonly called a thinking disposition, is a tendency to engage with a problem in a certain way. For example, individuals of a particular thinking style may tend to invoke a certain, possibly pre-determined strategy when faced with problems. One thinking style, as defined by Jonathan Baron (1993), is actively open-minded thinking (AOT): the tendency to seek out evidence that goes against one's own beliefs. This is a search strategy that can be employed on many problems. Another possible type of thinking style is one that involves motivation rather than strategy; people with higher motivation to engage with a problem may apply existing strategies differently (possibly more successfully), or they may be more successful at discovering new strategies for a given problem (Newton & Roberts, 2005). Need for Cognition (NFC) (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, & Jarvis, 1996; Cacioppo, Petty, & Kao, 1984) is one such thinking style; it is defined as the tendency to seek out, engage in, and enjoy effortful cognition or reflective thinking. Items of the scale used for measuring NFC include, "I would prefer complex to simple problems" and "I prefer my life to be filled with puzzles that I must solve" (see Appendix A).

Fluid intelligence, on the other hand, is defined as a cognitive *ability* rather than cognitive motivation or strategy. One useful definition of fluid intelligence is as the ability to understand complex relationships, reason, and solve novel problems independently of prior knowledge (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Martinez, 2000). When placed besides a motivation-based thinking style like NFC, the two can be likened to any other combination of effort and natural ability. For example, a given marathon runner may be naturally gifted with lungs and muscles

well suited to long-distance running – giving her a natural ability at the task – while a second runner may not have a body that is as well-suited. The two runners may also differ in the motivation or effort they put forward at the task, such that perhaps the second runner makes up for his shortcomings by exerting a greater effort in training. Importantly, just because the first runner has a greater potential does not mean that she will fulfill that potential; this requires motivation. Spearman's psychometric *g* (Jensen, 1998; Spearman, 1904), the general factor that underlies all intelligence tests, is very closely related to fluid intelligence (Duncan, Burgess, & Emslie, 1995), and so fluid intelligence is commonly thought to be a good measure of cognitive ability[1].

Both cognitive ability and thinking styles have been found to relate to the ability to solve problems in a normative way. Stanovich and West (1998, p. 161), speaking of the heuristics and biases experiments that make up the literature, noted that "...although the average person in these experiments may well [commit errors], on each of these tasks, some people give the standard normative response." Importantly, in that article, Stanovich and West not only showed that both cognitive ability and thinking style affect performance on a variety of tasks, but that they contribute variance uniquely – that is, they each play a unique role, such that both are certainly important. In their Experiment 1, they examined Wason's selection task, a syllogistic reasoning task, a statistical reasoning task, and an argument evaluation task. They used measures of AOT, counterfactual thinking, absolutism, dogmatism, and paranormal beliefs to measure thinking styles, and measured cognitive ability with the Raven Advanced Progressive Matrices (Raven, 1962; Raven, Court, & Raven, 1977), a common measure of fluid intelligence (Carpenter, Just, & Shell,

[1] Due to the strong relationship between fluid intelligence and *g*, and given that several research groups use measures of fluid intelligence to stand in for *g*, I will henceforth use the term "cognitive ability" as a catch-all phrase to include both.

1990), in addition to an academic aptitude test (the SAT) – this, like fluid intelligence, also loads highly onto *g* (see, e.g., Carroll, 1993). They grouped these measures to form three composite scores: one for tasks, one for thinking styles, and one for cognitive ability. In a multiple regression, they found that both cognitive ability and thinking style contributed variance uniquely to performance on the cognitive tasks, though cognitive ability contributed more (19.8% vs. 2.9%). They replicated these results in a further experiment using a composite score for tasks measuring covariation judgment, hypothesis testing, outcome bias, and if-only thinking (Experiment 2). Thus, with these tests, they firmly established separate roles for cognitive ability and thinking style in several cases.

In later studies, researchers in rationality began to become interested in the other thinking style I mentioned: Need for Cognition. This thinking style is a natural choice because of its motivational role; as Cacioppo, Petty, Feinstein, and Jarvis say, "Need for cognition is thought to reflect a cognitive motivation rather than an intellectual ability…and thus it should be related to but nonredundant with intellectual ability" (1996, p. 215). Indeed, in that paper they summarise studies showing that it is typically modestly related to several indexes of intelligence, and recent research supports these findings (Fleischhauer et al., 2010). Further, NFC is correlated with a number of other measures that indicate its likelihood as a predictor for normative thinking, such as a person's tendency to formulate complex attributions, devote attention exclusively to a cognitive activity, and base judgments on empirical evidence (see Cacioppo et al., 1996). However, it has also been shown to be unique from cognitive ability: NFC accounts for significant unique variance in the recall of earlier arguments made, above and beyond cognitive ability (Cacioppo, Petty, Kao, & Rodriguez, 1986), NFC and cognitive ability each predict unique variance in a skill-

acquisition task measured by a complex aviation game (Day, Espejo, Kowollik, Boatman, & McEntire, 2007), and NFC predicts increased processing of complex messages when using an assessment that holds cognitive ability constant (See, Petty, & Evans, 2009).

Thus, some researchers recently set out to discover if there is a unique relationship between NFC and normativity. Overall, the support for such a unique role has been present, although not for every task studied. Kokis, Macpherson, Toplak, West, and Stanovich (2002) tested AOT and cognitive ability in children, much like Stanovich and West (1998) did for adults, also including measures of NFC and reflective thinking. The children were tested on problems of base-rate neglect, syllogistic reasoning, and probabilistic reasoning. Kokis et al. found, using several multiple regressions, that the tendency to give normative answers increased with both cognitive ability and NFC on the probabilistic problem, and each contributed variance uniquely. Similarly, Toplak and Stanovich (2002) reported that, in multiple regression analyses, cognitive ability and a composite measure of NFC and reflective thinking each predicted unique variance in participants' ability to correctly solve disjunctive reasoning problems. Some researchers have also found relationships suggestive of potentially unique roles although without using multiple regressions to determine if unique variance is predicted. For example, Macpherson and Stanovich (2007) found that cognitive ability, NFC, AOT, and Superstitious Thinking were all significantly correlated with participants' ability to avoid belief bias when presented with syllogisms. Also, Stanovich and West (1999) found that participants who were more likely to give the normative response on a sunk cost problem were also more likely to have both higher cognitive ability and higher NFC.

Furthermore, in several cases researchers have reported that, even when cognitive ability is not related to the avoidance of bias, NFC sometimes is. Klaczynski and Lavallee (2005) reported that cognitive ability did not predict any additional variance above and beyond measures of epistemic regulation (including NFC) in avoiding bias in responding to goal-threatening (vs. non-goal-threatening) arguments. Additionally, Klaczynski, Gordon, and Fauth (1997) found that NFC was positively related to participants' ability to avoid bias in law of large numbers problems (wherein participants must evaluate the validity of evidence based on a small sample), while cognitive ability was not; Wedell (in press) found a similar pattern of results for avoiding conjunction errors. Stanovich and West (1999) reported that those participants who were more likely to give normative responses on the Wason selection task and Newcomb's problem were more likely to have higher NFC, while there was no significant difference between mean SAT scores for those choosing normative vs. non-normative responses. Finally, Klaczynski and Robinson (2000) found that those high in NFC tended to avoid myside bias in evaluating scientific experiments, while there was no relation between myside bias and cognitive ability. (However, Macpherson and Stanovich (2007) failed to replicate the myside bias effect and Stanovich and West (2007) only found an association between NFC and myside bias avoidance in one of several cases.) Overall, it is clear that NFC has a unique influence on normativity across a number of problems from the classic heuristics and biases literature.

1.3. Dual and tri-process theories: NFC's role in cognition

Once researchers established that thinking styles, and NFC in particular, play a crucial role in human rationality, they began to consider their inclusion in models of cognition. Typically, the model considered by heuristics and biases researchers is a

dual process model (for reviews, see J. S. B. T. Evans, 2008; Stanovich, Toplak, & West, 2008). This kind of model posits two types of processing: Type 1 processing is unconscious, automatic, and fast, and Type 2 processing is conscious, deliberative, and slow. Type 1 processes are generally described as running in parallel while Type 2 processes are sequential/serial, requiring one step to finish before another can begin – hence the differences in speed. Examples of Type 1 processing are emotions (in the case of decision-making, consider disgust and fear) and implicit associations (such as an association between birds and flight). Type 2 processing, meanwhile, is posited to *override* Type 1 processing in certain cases, stepping in to perform its slow, analytic functioning when called to do so. Much of this Type 2 processing is deemed to be hypothetical thinking (J. S. B. T. Evans, 2007), in which the thinker imagines several possible hypothetical situations while keeping them distinct from one another (and from any actual situation in reality); this is also referred to as cognitive simulation. Stanovich et al. also describe a second kind of Type 2 thinking, which they call serial associative cognition, inspired by J. St. B. T. Evans and Over's (2004) discovery that people tend to verbally discuss the cards they will choose on the Wason selection task, but only in terms of verification; people also only tend to look at the cards they will choose (J. S. B. T. Evans, 2007). In serial associative cognition, the thinker proceeds in thinking about a series of things – e.g., the vowel and (incorrect) even number card in the Wason selection task – but, crucially, does not imagine alternatives. In this example, the person thinks solely of the potential universe where the rule given is true ("if there is a vowel on one side, there is an even number on the other side"), and then serially considers cards to verify this rule. She does not consider alternate possible universes, where the rule is not true. This is an example of Type 1 and Type 2 processing working in tandem; Type 1 processing produces a heuristic

that says to treat the rule as true, and Type 2 processing serially examines the appropriate cards, perhaps also using Type 1 processing while examining each individual card. Because the thinker focuses on only a single model (where the rule is true), Stanovich et al. dub this error a focal bias. They claim that it is a potential error in problems of confirmation bias such as the Wason selection task, and the central error behind the framing effect (Tversky & Kahneman, 1981), in which participants focus only on the problem as framed (e.g., "200 out of 600 people will be saved for certain" as opposed to "400 out of 600 people will die for certain") and choose differently depending on which frame is presented. J. St. B. T. Evans (2007) agrees that such processing is responsible for a number of errors; he refers to it as the *fundamental analytic bias* due to its pervasiveness.

Altogether, researchers often argue that non-normative responses are indications of failures to invoke the useful, simulation-based Type 2 processing. Stanovich et al. (2008) outline a taxonomy of thinking errors to explain the possibilities in more detail (see Figure 1.1), which they claim sometimes involves this particular error but other times does not. Specifically, they first concede that possible failures to use Type 2 processing appropriately may be due to cognitive miserliness – the tendency to avoid spending cognitive effort if possible – causing a failure to override Type 1 processing, or causing the employment of Type 2 processing only in serial associative cognition. They also note that people may experience a simple failure to sustain the simulation-based Type 2 processing needed (this is often called decoupling, in reference to removing oneself from the actual world and simulating hypothetical worlds). Finally, people may suffer from problems with what they call *mindware* (coined by Perkins, 1995), which is the collection of knowledge, rules, and strategies a person can employ. It may be that there is a gap in this mindware (e.g., the

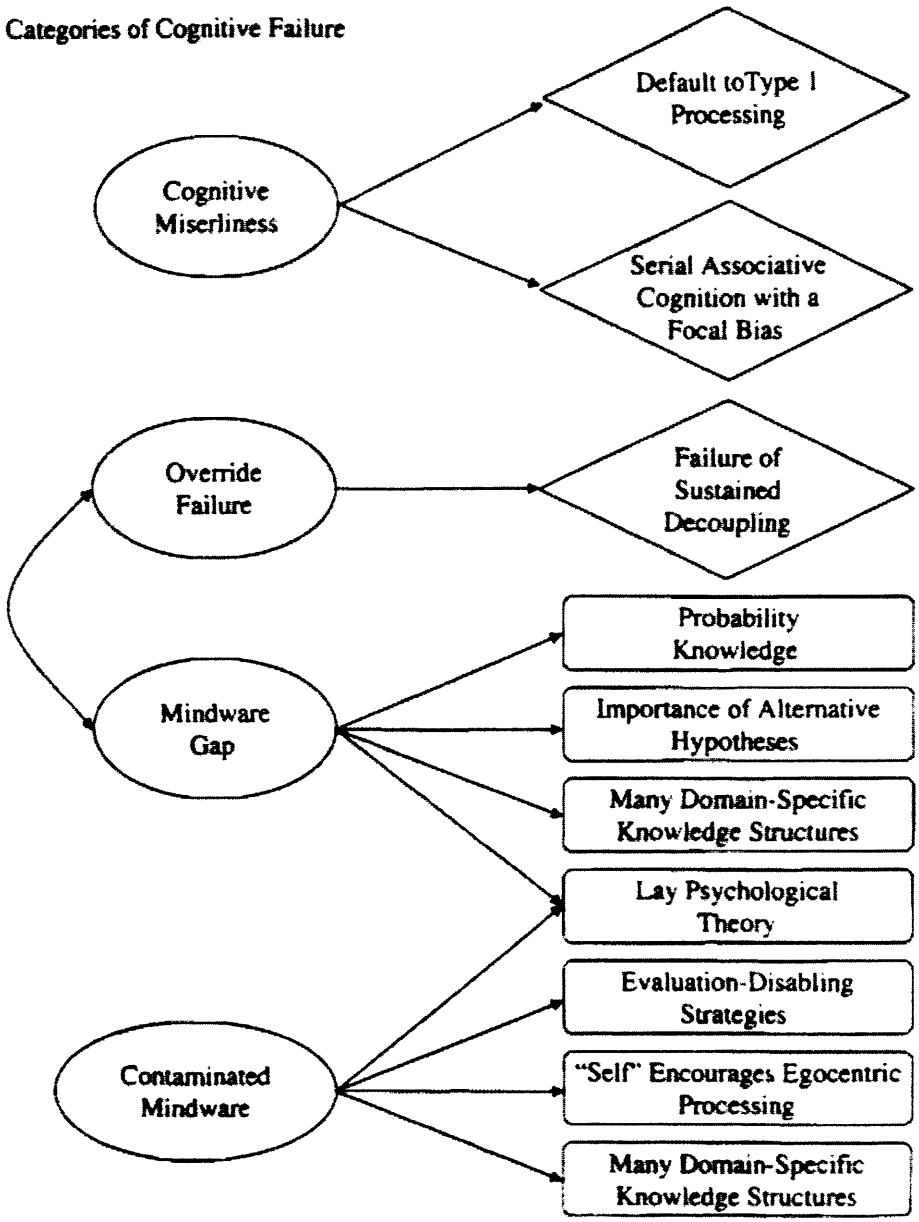


Figure 1.1. A preliminary taxonomy of thinking errors. Figure and caption reprinted from "The development of rational thought: A taxonomy of heuristics and biases," by K. E. Stanovich, M. E. Toplak, and R. F. West, 2008, *Advances in Child Development and Behavior, 35,* p. 260. Copyright 2008 by Elsevier.

person simply doesn't know a needed rule) or the mindware may be "contaminated" (e.g., people are egocentric and this may result in the myside bias). Of course, as J. St. B. T. Evans (2007) points out, the use of Type 2 (simulation) processing is certainly not a guarantee of a normative response and can sometimes cause a non-normative

response instead. This, he notes, may explain the controversial findings that people are more satisfied with complex decisions when these are made unconsciously, i.e., after spending time consciously considering unrelated things (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006). Still, in the scope of the literature, such Type 2 errors are dwarfed in number by the errors that researchers claim can be overcome by employing override properly in order to invoke Type 2 processing (along with possessing the appropriate mindware). NFC, and thinking styles in general, are proposed to account for just this ability to invoke the right kind of Type 2 processing at the right times.

Thus, researchers began to incorporate thinking styles into the dual process model (Stanovich, 2008; Stanovich & Stanovich, 2010), turning it into a model with three types of processes; they referred to these as the autonomous mind, the algorithmic mind, and the reflective mind (see Figure 1.2). The autonomous mind encompasses Type 1 processes, while Type 2 thinking is divided into the algorithmic mind, which determines fluid intelligence and therefore the efficiency and capability of completing a cognitively demanding task, and the reflective mind, which encompasses a person's beliefs, goals, and thinking styles. Stanovich and Stanovich explain that the reflective mind is responsible for initiating calls to the algorithmic mind: calls to override the autonomous mind, calls to begin mental simulation, and calls to change a current strategy (such as serial associative cognition). Meanwhile, the algorithmic mind is responsible for the actual simulation, as well as sustained attention such as that required by serial associative cognition. Uncontaminated mindware also continues to be required to produce rational responses, and can be linked to each of the three minds; the autonomous mind accesses evolutionary instincts and processes practiced to the point of autonomy, the algorithmic mind may

access various strategies and rules for sequencing events, and the reflective mind accesses beliefs, goals, and knowledge (particularly, I would argue, of certain strategies).

It is in the algorithmic and reflective minds that individual differences occur in order to produce the greater rationality already discussed. NFC plays a crucial role here, as it corresponds with the tendency to engage the algorithmic mind, and to do so

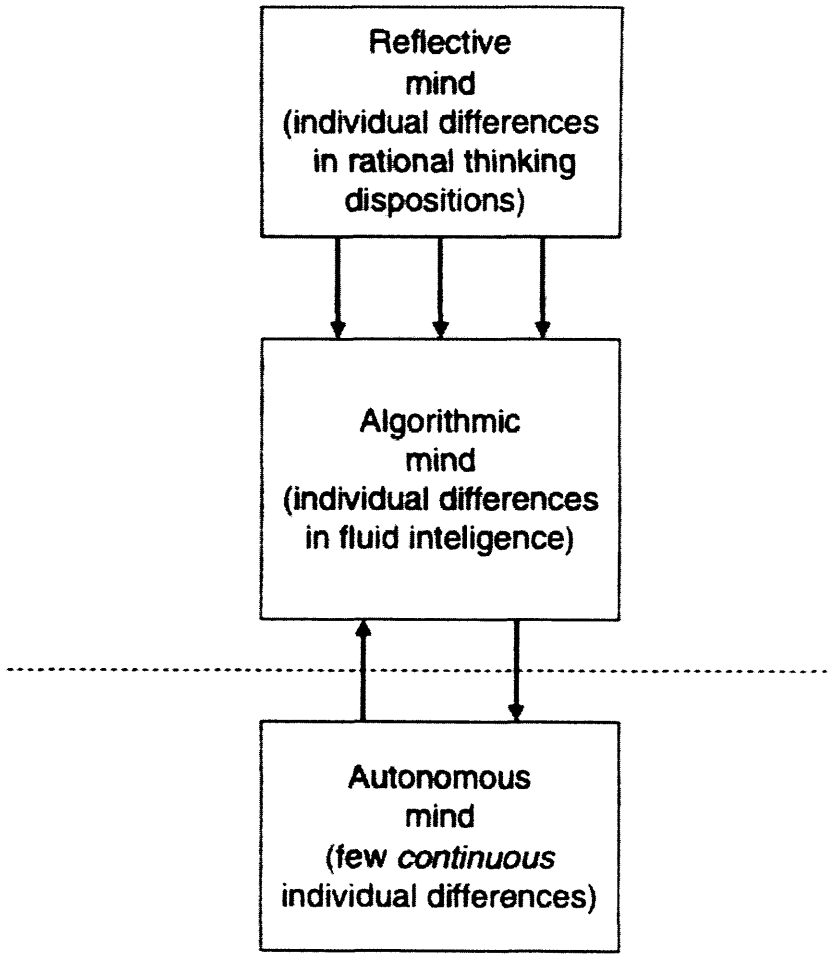


Figure 1.2. Individual differences in the tripartite structure. Figure and caption reprinted from *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (p. 211), by D. Preiss and R. J. Sternberg (Eds.), 2010, New York: Springer. Copyright 2010 by Springer. Reprinted with permission.

without merely resorting to the focal bias of serial associative cognition. Presumably, individuals high in NFC are devoted to assessing the problems with which they are presented in order to determine if analytic thought is needed, and are highly likely to carefully consider alternatives.

I will note that there is but one minor flaw in the tripartite model I have just presented: It assumes that the reflective mind falls within the remit of Type 2 processing. But why is it that the thought or feeling that one must override autonomous or heuristic processing is not considered a heuristic in itself? J. St. B. T. Evans (2006) clearly suggested that this decision to override is made by the autonomous mind, and Thompson (2009) concurs, describing a *feeling of rightness*; the weaker this feeling, she argues, the more one is likely to invoke Type 2 processes. Franssens and De Neys (2009) provide evidence for the automaticity of this feeling, showing that, even when performing problems under cognitive load, participants are able to detect a need to employ Type 2 processing. That the trigger to engage in Type 2 processing occurs while under load suggests it is itself not computationally intensive. NFC, then, may involve a fairly simple associative process. This, I would argue, would better fall under the heading of mindware, as would knowledge, beliefs, and goals, rendering the need for a separate reflective mind obsolete.

To summarise, the research on heuristics and biases tasks suggested a dual process model, for which considerable evidence has now been gathered (e.g., J. S. B. T. Evans & Curtis-Holmes, 2005; Franssens & De Neys, 2009; Roberts & Newton, 2002). The addition of individual differences as unique contributors to rationality suggested, to some, a tripartite model (Stanovich & Stanovich, 2010), although I argue that a more appropriate model is a dual process model in which thinking styles

are regarded as mindware. NFC, in this model, is a heuristic process that represents the individual's tendency to engage the algorithmic mind.

However, NFC's presence as one of many heuristic processes does not make it unimportant. It is unique in that it may lead to the development of both other important pieces of mindware and to the enhancing of the algorithmic mind. First, it may be related to building uncontaminated mindware: the more one tends to think deeply about problems, the more likely one may be to develop strategies for individual problems that can be called during appropriate future situations. Second, given that fluid intelligence can be improved via training one's working memory capacity (Jaeggi et al., 2008), i.e. via training the algorithmic mind, and NFC reflects a tendency to use this mind more often, NFC may encourage the very training that leads to greater fluid intelligence. Thus, NFC may be a crucial component not just in rationality itself, but in learning rationality.

1.4. Teaching rationality

Naturally, the largest reason researchers have become interested in studying thinking styles separately from cognitive ability is to determine the nature of human rationality – to answer the question of what makes us rational. This, of course, has applications for both creating rational computer programs and for creating other rational people; i.e., for teaching students. Although recent findings have indicated that it is possible to increase fluid intelligence with practice (Jaeggi et al., 2008), and the teaching of specific rules can also help overcome biases (Nisbett, Fong, Lehman, & Cheng, 1987), thinking styles are, as argued above, a separate avenue that should be explored due to their unique role. Further, Jonathan Baron (1985) has suggested that thinking styles are both simpler and easier to manipulate than cognitive ability. Baron hypothesizes that, rather than requiring hours of separate practice on working

memory, as Jaeggi et al. asked of their participants, or even brief training courses, as used by Nisbett et al., thinking styles can be manipulated by simple instruction: by, for example, telling someone to spend more time on problems before giving up. Such instructions could be threaded throughout any teaching program, and may provide crucial support for overcoming the biases discussed, in addition to the potential benefit of providing motivation to train one's working memory capacity and thus improve the intelligence of the algorithmic mind.

A further benefit of the inclusion of thinking styles in our conceptualisation of rationality is that they are much faster and easier to test than intelligence, requiring only a simple questionnaire. Thus, knowing now that they play a role in rationality, not only should they be included as part of typical IQ testing (Stanovich, 2009; Stanovich & Stanovich, 2010), but they might be used as quick measures in cases where full testing isn't practical.

1.5. Potential drawbacks: Is deep thinking always rational?

In line with my earlier discussion of local vs. global normativity, I must highlight that there is often debate over which solution is locally normative (e.g., Hahn & Warren, 2009; Oaksford & Chater, 1998a), and whether, for a given problem, local normativity should give way to global normativity (Oaksford & Chater, 1998b). In our society's preparation of our children to serve functional and useful roles, it can be difficult to determine which to choose: Do we teach them to use the locally normative strategies, possibly at the expense of useful globally normative strategies? Might it, at some times, be a globally-irrational idea to teach local-rationality? If so, the teaching of NFC may have potential drawbacks; although promoting the use of locally normative strategies on many problems, these may not be ideal on a global scale. I do not hope to be able to answer, in this thesis, whether or not these locally

normative strategies are a bad idea overall, as determining such would require knowing what constitutes a good global strategy – which is itself an exceedingly complex and difficult problem to solve. I simply note that this consideration should be kept in mind.

A more tractable question is whether NFC, while providing enhancements in rationality for some problems, may detract from performance on other problems. Few biases have been associated with conscious deliberation, save the tendency to report more satisfaction from deliberating "consciously" on simple choices and deliberating "unconsciously" on complex choices (Dijksterhuis et al., 2006). However, there is one glaring omission in the literature: No one has yet examined NFC's relationship to the expenditure of time. Given that Type 2 processes are theorized to be slow, and given the explicit mention of time within the NFC scale ("I find satisfaction in deliberating hard *and for long hours*," emphasis mine), there is certainly a time cost associated with deep thinking, and this cost may not always be warranted. Researchers have reported that NFC is related to a self-reported willingness to spend time (Verplanken, Hazenberg, & Palenéwen, 1992), to a desire to gather information (Berzonsky & Sullivan, 1992), to a greater store of accrued information (Inman, McAlister, & Hoyer, 1990), and to the valuing of web sites' information characteristics (Kaynar & Amichai-Hamburger, 2008). Further, Curşeu (2006) found that participants high in NFC took longer to respond on short-answer decision-making tasks. However, no one has yet examined the effect of NFC during a dynamic, real-time information-gathering task, and the potential role of cognitive ability has also been ignored. Since NFC and cognitive ability are related, both must be examined simultaneously in order to determine whether one or both are contributing uniquely.

In this thesis, I undertake the beginning of such an examination. The problem I examine is one of information gathering: The thinker must gather information about several options before deciding which is the best. I refer to this problem as one of *sampling-based choice,* as it involves sampling information in order to make a choice.

1.6. Sampling-based choice: A brief introduction

The problem of sampling-based choice is not one typically considered in the heuristics and biases literature; in fact, it has its own rather large and entirely separate literature. My purpose here is not to review that literature in its entirety, but rather to provide a basic outline of what the problem is, how theorists have attempted to model it, and what might be considered a normative response.

In its simplest form, the problem is one of statistical inference. Imagine a population about which a person wishes to learn – say, tree frogs. In the population, there is a mean and standard deviation for the hopping-length of the tree frog, but it would be impractical to gather every frog existing to measure this. We do so via sampling; we take a sample of frogs, measure their hops, and use the sample average to make an estimation about what the population average is. By the law of large numbers (Bernoulli, 1713), the larger our sample is, the more accurate our prediction. This is the essential rule of gathering information.

Once we have the information, we may wish to compare it to other information. Say we want to compare several things – e.g., several different species of frog, including the tree frog. We must undergo the same process as with the tree frog, but for each species, and use our knowledge of the mean and standard deviations of the samples to determine things about them, like if they are different (in statistics, this is done with a *t* test or an ANOVA). A large amount of empirical research has shown that people are good "intuitive statisticians" when it comes to this kind of problem

(see Sedlmeier & Gigerenzer, 1997, for a review). Although intuitions are not always accurate (e.g., Tversky & Kahneman, 1971), people are sensitive to the law of large numbers during such frequency-estimation tasks as the sampling-based choice problem.

The way that theorists typically model sampling-based choice involves a sequentially updated store of the information seen thus far (e.g., Einhorn & Hogarth, 1981), such that evidence accumulates for all options, in either an integrated way (as with *random walk* models) or separately (as with *counter* models) (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Busemeyer & Townsend, 1993; Pleskac & Busemeyer, 2009; Ratcliff & Starns, 2009; Smith & Vickers, 2009). This accumulated information is always compared to a threshold, such that the decision maker stops to make her choice once the strength of the evidence exceeds the threshold. In considering our tree frog example above, the way a modern statistician would compare the hop-lengths of the tree frog to the ground frog would be to sample from each population, then examine the mean difference in hop-lengths and the standard error of this difference. If the difference is large enough, and the standard error small enough, she could say that there is less than a 5% chance of obtaining such a large difference were the two populations actually equal. This is the process of a *t* test; our statistician's evidence strength is determined by the mean and standard error (which together make up the *t* statistic), and her threshold is 5% (or 95%, depending on how you look at it). Although people might not actually be literally performing *t* tests in their heads (though it has been suggested, see Kelley, 1967), the fact that they are good intuitive statisticians on such problems, and that numerous supported models suggest a similar process, together indicate that they are doing something close.

Importantly, gathering information always takes both time and effort, in exchange for the knowledge it produces. This is known as the speed-accuracy trade-off, a phenomenon that has been well known for some time (Garrett, 1922; Hick, 1952; Schouten & Bekker, 1967; Wickelgren, 1977; Woodworth, 1899). Thus, the normative strategy is to optimise based on the utility one gains from large-sample accuracy vs. the utility lost from spending time. Naturally, this strategy will depend on the goals of the individual; perhaps a given person values accuracy a great deal and is willing to make this trade-off for time. In particular, I hypothesize that individuals high in NFC may display such a preference for accuracy. However, of course, if strong individual preferences persist in the face of a different experimenter-defined set of utilities, we might then call the individual's strategy non-normative.

It is also worthwhile to note that in many cases the cost of accuracy can be considered quite steep. First of all, drawing sample items often requires storing them in memory; this can be taxing and, over time, requires an increasingly great effort. Hertwig and Pleskac (2008) describe how this memory problem is compounded by the simplicity of the small sample. They mathematically prove that an experienced difference between two samples is always larger for small samples than for large samples (the *amplification effect*). For example, they consider a participant drawing from two payout decks. Deck A has cards that offer $32 with probability .1, $0 otherwise, and deck B offers a payout of $3 for sure. When the participant draws from deck A, a small sample means he is likely to miss the low-probability event of seeing $32. Thus the expected difference may seem to him to be $3 ($0 from deck A and $3 from deck B), when in fact it is $0.20. The large difference in the case of the small sample ($3 rather than only $0.20), Hertwig and Pleskac argue, makes choosing from smaller samples simpler and easier.

Further, the speed-accuracy trade-off is described by a curve of diminishing returns. For example, Hertwig and Pleskac (2008) report that, using the natural mean heuristic – a reasonable strategy that we might expect people to employ – a sample as small as 1 results in a 60% chance of determining the deck with the better expected value. Sampling to 5 gives a 78% chance, to 10 an 84% chance, and to 20 an 88% chance. Other strategies provide a similar curve. Thus it always requires ever-increasing effort to gain a small percentage increase. A sampler who wishes great accuracy, therefore, pays a premium in time. A normative sampler must value accuracy very highly in order to gain from paying this price.

Importantly, there is evidence that people who deliberate for longer may in fact be less satisfied. People report less satisfaction when deliberating consciously vs. unconsciously on a complex choice (Dijksterhuis et al., 2006), and maximizers – people who search continuously through available options in order to find the "best" – also report lower satisfaction when faced with consumer choices (Schwartz et al., 2002). Perhaps a focus on accuracy is not always part of a simple normative strategy; the high cost of accuracy described above may not be estimated properly. Thus there is certainly a potential for NFC to be related to non-normative behaviour, even in a participant-defined sense.

There are, then, two important reasons to examine the relationship of NFC to time on this task. First, if individuals behave according to certain preferences (such as a preference for accuracy), it is useful to know this in some cases. For example, if a person is hiring employees who need to make this kind of sampling-based choice, she may want to know whether they may focus on accuracy over speed, so that she can choose to hire those who will help her optimise her own goals. Second, NFC may be related to a failure to perform in an experimenter-defined normative way, with high

NFC individuals spending extra time without concurrent benefits in accuracy. A major focus on accuracy, if it exists among those high in NFC, may also be indicative of a potential failure in participant-defined normativity.

1.7. Summary and experiment plan

Researchers have conceptualised rationality or normativity in terms of optimising a goal set, either for a set of problem-specific, local goals, or a broader set of global goals (although the empirical focus has been on the former). Need for cognition (NFC), or the tendency to enjoy expending cognitive effort, has been found to relate uniquely to normative performance on a number of traditional heuristics and biases tasks, suggesting that it plays an important role in rationality separately from cognitive ability. I suggest that it is incorporated in cognition as a fast, associative (Type 1) process that monitors problems for the need to call the slow, serial Type 2 processing, of the type that involves mental simulation or consideration of alternatives.

My experimental plan focuses on exploring whether NFC is ever related to non-normative performance. I do this by examining a sampling-based choice problem, for which there is a speed-accuracy trade-off; I argue that it is informative to understand whether there is a simple relationship between NFC and spending more time. In the experiments that follow, I consider first whether this relationship exists. After finding in Experiments I and II that it does, I then consider for what reasons it may exist, and whether these reasons render the performance normative. Given the current evidence that deliberating or searching for longer does not always result in greater satisfaction, then a focus on accuracy – the major hypothesized cause – may not always be normative, either in an experimenter- or participant-defined sense. However, I also explore whether the effect is driven by a simple lack of ability, an

overarching tendency to think about the task (e.g., in order to improve, or to find patterns in the data), a desire to appear in a positive light, distraction due to boredom, or a simple heuristic to spend time, and I discuss whether and how each of these possibilities may be considered normative. I have included a complete discussion of these hypotheses and the experiments designed to test them in the discussion of Experiment I.

1.7.1. Analysis overview: Assumptions and analysis procedures used for all experiments

In order to reduce the complexity and difficulty of both reading individual analyses and comparing analyses across experiments, I summarise here the procedures I used in all cases. First, I always examined correlations pairwise and assessed and removed outliers by the Jackknifed Mahalanobis distance method (see, e.g., Robinson, Cox, & Odom, 2005), which is identical to the Mahalanobis distance method (Mahalanobis, 1936) but ignores the given observation when calculating statistics about the data set, thus reducing the risk of undue influence if the observation is an outlier.

Second, whenever employing a statistical test relying on assumptions of normality, I used the Kolmolgorov-Smirnov test for samples below 50 in size, and the Shapiro-Wilk test for samples of size 50 or above; this is because, as an exact test, the Kolmolgorov-Smirnov test is more suitable for small samples. Further, whenever I discovered violations of assumptions, including normality and homogeneity of variance, I either a) transformed the data, where suitable (these instances are noted in the text), or b) employed, where possible, nonparametric versions of the tests, such as the Kruskal-Wallis nonparametric ANOVA and the Mann-Whitney U test. Because nonparametric tests were required in making between-experiment comparisons in

some cases, I chose to use them in all cases for the sake of consistency. Results in these cases were not significantly different from those obtained with parametric tests. An alpha level of .05 was used for all statistical tests.

Finally, in some cases, due to violations of assumptions, it was necessary to log-transform scores, in particular recordings of time taken to complete tasks. Again, for the sake of consistency, I thus report all time-based scores as log-transformed, unless noted otherwise.

2. Experiment I: A long search for information

Experiment I explored people's preferences for sampling evidence before making a choice. More specifically, it investigated whether there would be a relationship between Need for Cognition (NFC) and the amount of time spent sampling, one that could not be explained by cognitive ability alone.

According to previous research on NFC, it appears to be related to evidence search; for example, there is a significant positive correlation between NFC and people's self-reported tendencies to seek out and scrutinize information when making decisions (Berzonsky & Sullivan, 1992), and their accumulated store of consumer information (Inman et al., 1990). Additionally, Curşeu (2006) found positive correlations between NFC and time spent on typical short-answer decision making tasks, and Verplanken, Hazenberg, and Palenéwen (1992) found that individuals high in NFC were more likely to indicate a willingness to spend time gathering information than those low in NFC. Finally, Baugh and Mason (1986) found that NFC was negatively correlated with time estimation of an anagram task; this suggests that individuals high in NFC may search for longer simply because it does not seem to them as if as much time were passing.

However, no research has yet been done to determine just how much searching high-NFC individuals (henceforth, high NFCs) will do in a real-time, dynamic decision task, one in which they have full control over the amount of information they gather (and therefore the time they spend gathering). Furthermore, none of the above research rules out the possibility that cognitive ability, which is often correlated with NFC (Cacioppo et al., 1996; Day et al., 2007; K. E. Stanovich & R. F. West, 1998; Wedell, in press; West et al., 2008), is driving the effect. If NFC contributes uniquely to this prolonged information search, then this would indicate

that there is something important about the *attitude* that people adopt in making decisions, represented by NFC, rather than simply their ability to complete the task effectively. Essentially, NFC has the potential to represent the effort put forward toward a cognitive goal, as opposed to cognitive ability, which only represents the capacity to remember and manipulate information. It is therefore important to distinguish the separate effects of the two measures.

This experiment presented the participants with a simple, two-alternative decision to make: Gather information about two potential job candidates in order to determine which is better. To simplify the task and remove any variance due to complexity of presented information, participants saw only positive or negative information during the search (represented by smiley faces and frowny faces, respectively[2]). These were presented on a computer screen, divided for the two candidates, and the participant was able to hold down a button in order to view information (faces) appearing for each of the candidates, releasing the button when ready to choose. I predicted that amount of time spent on the task, which in this case is equivalent to the amount of information gathered, would be correlated with NFC, with high NFCs spending more time. I also predicted that this relationship would be independent of cognitive ability.

There may be several potential reasons for this predicted relationship. In particular, I predicted that, if NFC were related to time spent, the reasons might include 1) a focus on accuracy over speed, 2) expending extra time thinking about the

[2] The choice to use faces to display this information was based entirely on replicating the method of Fiedler and Kareev (2006). It should be noted that we might expect differences in performance if, say, the information were relayed verbally (e.g., using the words "positive" and "negative") or otherwise (e.g., using "+" or "–"). Past research has shown that those high in NFC tend to prefer verbal to visual information (Martin, Sherrard, & Wentzel, 2005; Sojka & Giese, 2001); however, the question of whether such a preference would lead them to sample more (e.g., due to enjoying the task more) or less (e.g., due to quicker processing of the stimuli that they like better), I leave to future research.

task (for example, in order to improve at it, or to search for patterns), 3) spending more time in order to please the experimenter (i.e., due to social desirability bias), 4) being distracted due to boredom, or 5) engaging a heuristic or rule of thumb that time should be spent, for no particular reason. The NFC-time relationship may also be accompanied by a relationship to accuracy and also, therefore, a relationship to confidence (as one who often gives a correct answer will be likely to have more confidence in her answers in general). Because my focus is on the speed-accuracy trade-off, the primary variables I measure are time, accuracy, and confidence. However, in order to help determine potential reasons for the NFC-time relationship, I also include study of participants' ability to improve with time (which would indicate a potentially normative reason for thinking deeply at first – if it saves time in the long run), and their reported focus on accuracy or efficiency.

2.1. Method

2.1.1. Participants

31 adults (21 female and 10 male) were recruited from the Cardiff University Human Participants Panel, with median age 23. They were paid £6 for their involvement.

2.1.2. Procedure

Participants were always tested alone, with the experimenter sitting in the next room. They entered and were seated at a computer, which provided both the bulk of the instructions and the tasks. They first completed the choice task; this was followed by the cognitive ability measure and then the questionnaire, which included portions

for NFC and social desirability (SDR). All questions in the questionnaire were randomly intermixed.

2.1.3. Choice task

Following the method of Fiedler and Kareev (2006), the task was a two-alternative choice task. Participants read a cover story that asked them to fill employee positions by choosing the superior of two candidates; each choice, they were told, would be based on a sample of evidence (see Appendix B).

Instructions asked participants to imagine that they were members of a personnel department whose job it is to choose whom to hire. Employee candidates had already taken part in an extensive testing process, where they were judged by a very large number of judges. Each judge had rated each candidate as either good (represented by a smiley face: ☺) or bad (represented by a frowny face: ☹). The instructions clearly stated that there were so many judges making judgments that it would not be possible to see them all. Therefore, participants needed to take a sample of judgments for two candidates at a time. Their ultimate goal was to decide from the sample which candidate was better, and to do so without taking too much time.

Equal emphasis was placed on the importance of accuracy and efficiency. The program intermixed words and phrases relating to accuracy (A) and efficiency (E) in the instructions in order to counteract any possible order effects (e.g., "Both accuracy and efficiency are important in working on this task. This means that, within the time available, you should compare as many pairs as possible, but at the same time try not to make any mistakes."). Phrases were always in an AEEA or EAAE format, which was randomly determined per participant. The word "efficiency" was used instead of "speed" to discourage moving through trials without gathering any evidence. I.e., some minimum level of required accuracy was implied.

Participants viewed a computer screen that had been divided into two halves; the left was labeled "Candidate A" and coloured turquoise; the right was labeled "Candidate B" and coloured orange. In order to take a sample of judgments, the participant clicked with the mouse and held down a button onscreen. During the button press, judgments (smileys or frownies) appeared for 500 ms each at random locations within the two coloured areas, with a 500 ms interval between faces. When participants released the button, the judgments stopped appearing, and the screen presented three options: Choose Candidate A, Choose Candidate B, or No Choice. Participants also rated their confidence in the choice they had just made on a 21-point scale, which was done using a sliding scale onscreen. After clicking to move on, a two second interval occurred, during which the screen was blank. The time participants took to complete the entire block was recorded in seconds.

For each trial, the program chose the better candidate randomly. During sampling, it drew judgments for each candidate randomly (with replacement) from one of two possible universes. A "weak evidence" universe was comprised of 55% positive and 45% negative judgments for the better candidate, and 45% positive and 55% negative for the worse. A "strong evidence" universe had 70% positive, 30% negative vs. 30% positive, 70% negative. All participants saw 10 trials from each universe, randomly intermixed.

At the end of the choice task, participants rated 1) the amount they focused on either accuracy or efficiency during the task (AE Focus – Task), and 2) how they tend to split their focus between accuracy and efficiency in general (AE Focus – General). The ratings were made on a sliding scale; the participant used the mouse to drag a button anywhere along a line between "100% on Accuracy" and "100% on Efficiency." Scores for this measure were coded on a 100-point scale such that a low

score (<50) indicated a focus on Accuracy while a high score (>50) indicated a focus on Efficiency. The button started in the middle of the scale (at 50).

2.1.4. Materials

2.1.4.1. Cognitive ability measure: Raven's Progressive Matrices

Participants completed a subset of problems from the Raven Advanced Progressive Matrices (Raven, 1962, hereafter referred to as the Raven matrices). Following the example of Stanovich and West (1998), I excluded the 11 easiest problems – at which performance is near ceiling for university-educated adults – and the 6 most difficult – at which performance is floored (Carpenter et al., 1990; Raven et al., 1977). This left 19 problems for participants to solve. Participants were given 16 minutes to complete these problems.

The Raven matrices problems consist entirely of pattern-finding tasks, wherein participants must identify which symbol fits two patterns: one given in rows of example symbols and the other in columns, presented in a matrix. The matrices are considered to be a good measure of analytic or fluid intelligence (Carpenter et al., 1990). Scores on this test were obtained by subtracting the number of incorrect answers from the number of correct, with no penalty for unanswered questions. The mean score was -1.93 (SD = 8.35). Cronbach's α for this measure was .87.

2.1.4.2. Need for Cognition (NFC) scale

I also asked participants to complete the 18-item short form Need for Cognition scale published by Cacioppo, Petty, and Kao (1984) (see Appendix A). Sample items are: "I prefer my life to be filled with puzzles that I must solve," and "I only think as hard as I have to" (reverse scored). Participants rated each item on a 7-point Likert scale from 1 ("Disagree Strongly") to 7 ("Agree Strongly"). Scores were

obtained by summing all responses; the mean score was 86.39 (*SD* = 13.34). High scores indicate an enjoyment of cognitively effortful tasks; low scores indicate a preference for avoiding cognitive effort. Cronbach's α for this measure was .90.

2.1.4.3. Social Desirability (SDR) scale

Following the example of Kokis et al. (2002), participants completed a 5-item scale designed to assess their tendency toward socially desirable responding. Four of these items were taken from the impression management subscale of the Balanced Inventory of Desirable Responding (Paulhus, 1991): "I always obey rules even if I probably won't get caught"; "I sometimes tell lies if I have to" (reverse scored); "There are times I have taken advantage of someone" (reverse scored), and "I have said something bad about a friend behind his or her back" (reverse scored). The final item, "I like everyone I meet," was taken from the Self-Esteem Inventory (Blascovich & Tomaka, 1991), which is known to correlate highly with social desirability. Participants rated each item on a 7-point Likert scale from 1 ("Disagree Strongly") to 7 ("Agree Strongly"). Continuous scoring (wherein, once items are reversed, all item scores are summed) was employed rather than dichotomous scoring (where extreme scores are counted more or are the only scores counted) at the recommendation of Stöber, Dette, and Musch (2002); the mean score was 17.75 (*SD* = 3.15). High scores indicate a tendency to exaggerate positive attributes – i.e., to present a positive impression to others – while low scores indicate a tendency for more honest self-portrayal. Cronbach's α for this measure was .44, a low score that may be due to the small number of items on the scale.

2.2. Results

One participant did not complete the Raven matrices and was excluded from analysis. A second was removed for scoring lower than 3 standard deviations from the mean on the NFC measure, and a third was removed for scoring -19 (all questions answered incorrectly) on the Raven matrices in addition to completing the choice task faster than all other participants; it was apparent that this participant failed to comply with instructions. Thus, 28 participants were entered into analysis. If, on a given trial, the participant viewed only a single judgment, I considered it an error in button-pressing; these trials were discarded from analysis. The median number of such errors per participant was 1, and the greatest number of errors made by a single participant was 4.

2.2.1. Descriptive statistics

I first measured the total time taken to complete all 20 trials in seconds. On average, participants took 683.76s (SD = 196.13) to complete the trials (11.40 minutes); logged scores had a mean of 2.79 (SD = 0.13). All further analyses are conducted on logged scores (see Chapter 1, section 1.7.1). Time taken was highly correlated with the participants' median sample sizes, $r = .95$, $p < .001$. Because these two measures were essentially interchangeable, I chose to examine time taken as opposed to sample size in order to better compare these results to past literature (Curşeu, 2006; Verplanken et al., 1992).

I determined each participant's accuracy by the number of correct candidate choices made minus the number of incorrect, ignoring trials on which the participant made no choice. Overall, participants achieved a mean score of 8.50 (SD = 3.69), getting an average of 12.46 (SD = 2.38) correct, 3.93 (SD = 2.00) incorrect, and rarely

deciding not to make a choice (*Median* = 2.50). The average score was significantly better than the chance score of 0, $t(27) = 13.11, p < .001$.

A third measure was participants' median confidence across all trials on which they neither made an error nor made no choice. On average, participants' median confidence was 12.57 (*SD* = 3.87) out of 21.

I also considered whether there might be a difference in individuals' ability to improve at the task, i.e., to attain more correct answers as they progressed. In order to examine this possibility, I assigned an improvement score to each participant by taking the point-biserial correlation between trial number (1-20) and accuracy on the trials (coded 1 for a correct choice, –1 for an incorrect choice, and 0 for no choice). If one's accuracy is improving, more 1s will be generated on later trials than earlier ones, and the result will be a positive correlation. The mean such score was –.01 (*SD* = .23); this suggests that some participants significantly improved at the task, while others became worse and most stayed about the same.

Finally, I examined participants' responses to the questions of how much they focused on accuracy or efficiency both during the task (AE Focus – Task) and in general (AE Focus – General). As expected, given that their instructions equally emphasised accuracy and efficiency, participants' average response for the task was 50.79 (*SD* = 20.96), which was not significantly different from 50 (the middle of the 100 point scale), $t(27) = 0.20, p = .84$, indicating participants in this sample generally reported equal focus on both accuracy and efficiency. When asked how much they valued accuracy and efficiency in general, their average score was 42.86 (*SD* = 18.28); this was significantly different from 50, $t(27) = -2.07, p = .048$, indicating an overall focus on accuracy.

2.2.2. Main analyses

Table 2.1
Experiment I: Intercorrelations among primary variables.

Variable	1	2	3	4	5	6	7	8
Ability Measures								
1. Cognitive Ability	–							
2. Need for Cognition	.36 §	–						
Choice Task								
3. Time Taken (log)	-.15	.42 *	–					
4. Accuracy	.03	-.21	-.11	–				
5. Median Confidence	-.08	-.17	-.38 §	-.02	–			
6. Improvement	.52 **	.14	.08	-.31	.26	–		
AE Focus								
7. AE Focus - Task	.11	-.25	-.38 *	-.08	.27	.31	–	
8. AE Focus - General	-.24	-.45 *	-.24	-.07	-.25	-.20	.32	–

Note. Outliers were assessed and removed in pairwise correlations by the Jackknifed Mahalanobis distance method (Mahalanobis, 1936; also see section 1.7.1). Time Taken refers to the total amount of time taken to complete all 20 trials. AE Focus refers to the participant's focus on either accuracy or efficiency; a low AE Focus score indicates a focus on accuracy and a high score indicates a focus on efficiency.

§ $p < .10$.
* $p < .05$.
** $p < .01$.

Table 2.1 displays intercorrelations between the primary variables in this study. There were several interesting trends. Most notably, NFC and Cognitive Ability (as measured by the Raven matrices) were marginally correlated, $r = .36$, $p = .07$, which is in accordance with the majority of literature on the Raven matrices (Day et al., 2007; Wedell, in press), although one null relationship between Raven performance and NFC has also been reported (Bors, Vigneau, & Lalande, 2006). Also of note was that AE Focus (Task) was significantly correlated with Time Taken, $r = -.38$, $p = .047$, such that more time taken was associated with a greater focus on accuracy. This correlation indicates that participants understood the accuracy-efficiency relationship to be primarily driven by time taken (in order to sample more). Also, NFC was significantly related to AE Focus (General), $r = -.45$, $p = .02$, such that those high in NFC reported a general preference for accuracy, as predicted.

Improvement was significantly related to Cognitive Ability, $r = .52$, $p = .007$, such that participants of higher cognitive ability were more likely to improve or maintain their performance. This relationship is in the predicted direction, given that those who score highly on the Raven matrices typically have a high working memory capacity (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004) and therefore more potential ability to remember faces as they become used to the task (i.e., after an initial period of adjustment). The discovery of such a predicted relationship suggests that the improvement measure is a valid one.

The relationship between Accuracy and Time Taken was nonexistent, $r = -.11$, $p = .59$, which may seem surprising given that accuracy and speed are commonly strongly associated (Garrett, 1922; Hick, 1952; Schouten & Bekker, 1967; Wickelgren, 1977; Woodworth, 1899), as predicted by the law of large numbers (Bernoulli, 1713). However, consider that the law of large numbers only says that a *fixed-size* large sample, compared to a fixed-size small sample, should perform better. In this task, participants did not receive samples of fixed size; they were constantly monitoring the sample and able to stop sampling whenever they wished. Thus, when early evidence was strong, they would likely stop sampling, and when early evidence was weak, they would continue. Indeed, several researchers have reported that the average time to make a choice decreases with strength of confidence in one option (e.g., Baranski & Petrusic, 1994; Henmon, 1911), and this trend appears in the current sample, with a marginal negative correlation between Time Taken and Median Confidence, $r = -.38$, $p = .054$. One very likely strategy the sampler might employ is to continue sampling until evidence has reached a threshold of strength, t, which results in accuracy a (L. Evans & Buehner, in press; Griffin & Tversky, 1992;

Hertwig & Pleskac, 2008). E.g., a person might like to sample until she is likely to be 80% correct, with the goal of reaching this accuracy level regardless of sample size. Although people tend to be overconfident with small samples and underconfident with large samples (Griffin & Tversky, 1992), it should still be possible to achieve similar accuracy in small sample and large sample situations. In this case, the result would naturally be a small or nonexistent correlation between Accuracy and Time Taken.

The full relationship between NFC, Cognitive Ability, and Time Taken, although unclear from simple correlational analysis, became clear with a regression analysis. The simultaneous regression of these two variables onto Time Taken revealed that both play a significant role in this task. Both Cognitive Ability and NFC were significant unique predictors of variance; NFC explained 19.3% of the variance uniquely, while Cognitive Ability explained 16.2% uniquely, both *p*s < .05 (see Table 2.2). Their standardised beta weights indicate that greater Cognitive Ability is associated with less Time Taken (–.424), which is consistent with Cognitive Ability's relationship to working memory capacity: Participants with high Cognitive Ability were better able to remember the faces they had already seen and therefore did not need to sample many more. Meanwhile, as predicted, greater NFC is associated with more Time Taken (.463), independently of Cognitive Ability.

Table 2.2
Experiment I: Simultaneous regression of NFC and Cognitive Ability onto Time Taken (log)

	β Weight	*t* Value	Unique variance explained	Partial *r*
Criterion variable: Time Taken (log)				
Cognitive Ability	–.424	–2.349*	.193	–.425
Need for Cognition	.463	2.568*	.162	.457
Overall regression: F = 4.60*				
Multiple *R* = .519				
Multiple *R*-squared = .269				

Note. *β* weights are standardised.
*$p < .05$.

One possibility is that this relationship between NFC and Time Taken is related to a focus on accuracy. Because there was a significant correlation between NFC and AE Focus (General), $r = -.45$, $p = .02$, even with Cognitive Ability removed, $r = -.40$, $p = .04$, I attempted to determine if it mediated the relationship between NFC and Time Taken (see Figure 2.1). To do this, I conducted the bootstrapping analysis described by Preacher and Hayes (2004; 2008), using 5000 bootstrap resamples and including Cognitive Ability as a covariate. The main benefit of this analysis over the more common causal steps analysis (R. M. Baron & Kenny, 1986) is that bootstrapping allows one to estimate the sampling distribution of the indirect effect (path *ab*; alternatively $c - c'$); it therefore does not rely on the assumption that this distribution is normal.

However, the test of the indirect effect was not significant, with a point estimate of .0002 and a 95% BCa (bias-corrected and accelerated confidence interval – see Efron, 1987) of −.0013 to .0018. This confidence interval includes zero, indicating that there may be no difference between the total effect of NFC on time taken (path *c*) and the direct effect (path *c'*), when AE Focus (General) is removed. Thus, there is no significant mediation by AE Focus (General).

In order to test for any potential roles of SDR, Median Confidence, or Improvement in this relationship, I also examined whether these variables were significantly correlated with NFC when Cognitive Ability was controlled; if so, they might also play mediator roles. However, there was no relationship between NFC and SDR, $r = .29$, $p = .14$, nor between NFC and Median Confidence, $r = .18$, $p = .44$, nor finally between NFC and Improvement, $r = .04$, $p = .85$. It is therefore apparent that there is no potential mediation involving these three variables. It is also clear that high

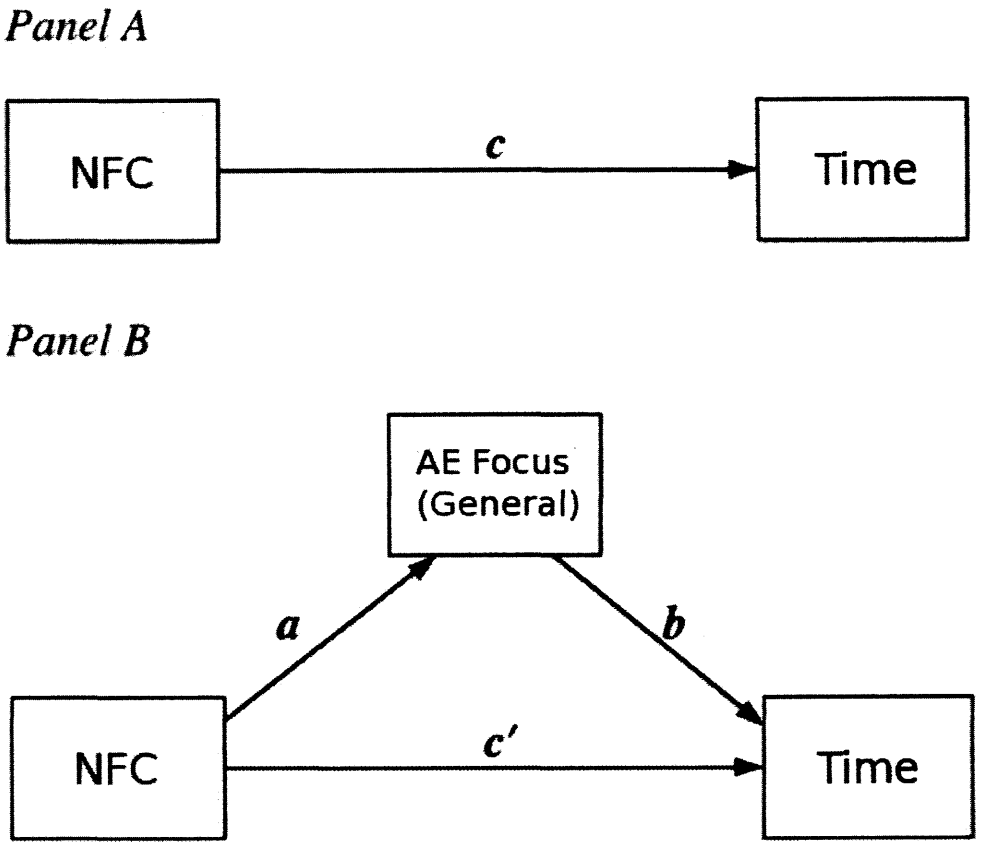


Figure 2.1, Experiment I. An illustration of the potential mediator role of AE Focus (General). Panel A displays the total effect of NFC on Time Taken (path *c*), while Panel B displays the direct effect (path *c'*) and the indirect effect (path *ab*) of NFC on Time Taken. I.e., NFC may affect Time Taken indirectly, through AE Focus (General).

NFCs in this sample, despite the extra time they spend, are not improving at the task any more than low NFCs, nor gaining any extra confidence.

Next, I tested whether the evidence universe a participant was viewing had any effect on the amount of time taken to sample. I had strong theoretical reasons to believe that the weak evidence universe would result in greater difficulty for the participants and therefore more time spent. I also sought to discover whether there might be an interaction between universe and NFC, such that, perhaps, the greater time spent by high NFCs occurred largely in one universe or the other. I thus

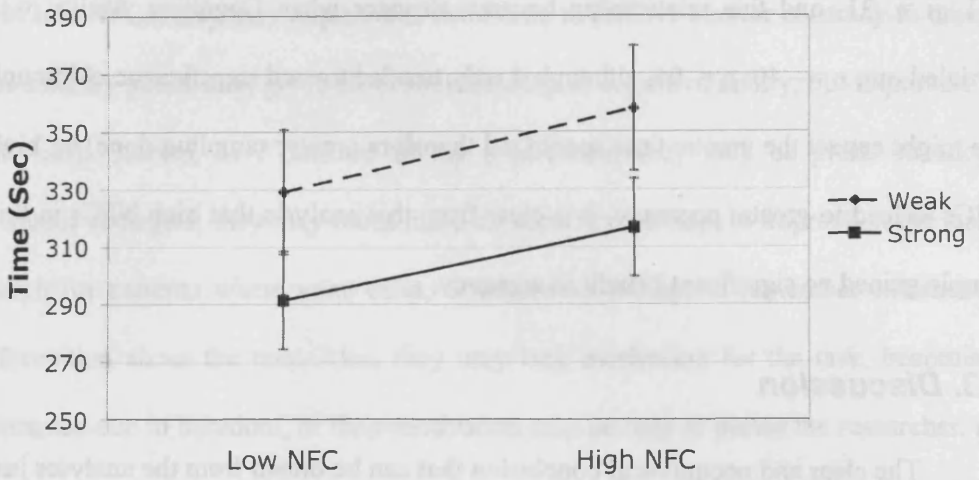


Figure 2.2, Experiment I. Mean time taken (in seconds) to complete all trials of either the weak or the strong evidence universe for participants low in NFC (left; estimated at $M - .5SD$) and high in NFC (right; estimated at $M + .5SD$). Bars are standard error.

performed a 2 (Universe: Weak or Strong) x NFC (measured)[3] ANOVA. Indeed, the predicted main effect of universe was present, $F(1, 26) = 12.04$, $p = .002$, $\eta_p^2 = .316$, such that participants spent more time in the weak universe ($M = 282.81$[4], $SE = 14.59$) than the strong ($M = 255.54$, $SE = 11.85$) (see Figure 2.2). However, there was no significant interaction, $F < 1$. Thus the apparent time-spending of the high NFCs did not differ based on the strength of the evidence seen. Note that there was no significant effect of NFC in this analysis, $F = 2.65$, $p = .12$, possibly because it did not partial out the effect of Cognitive Ability.

Finally, I tested the relationship between NFC and Accuracy. Regardless of high NFCs' significant focus on accuracy and their clear effort to spend time sampling, there was a slight negative relationship between NFC and Accuracy, $r =$

[3] Including NFC as a measured variable allows me to use the full continuous nature of the variable and therefore retain power. All analyses in this thesis including measured NFC were also performed using the median splits of NFC, with a similar pattern of results, unless noted.

[4] Note that it was not possible to perform the analysis on log-transformed time spent, as these values violated the assumption of normality; meanwhile, the untransformed scores did not. Thus the times reported are in seconds.

–.21, p = .31, and this relationship became stronger when Cognitive Ability was partialed out, r = –.40, p = .06, although it only trended toward significance. Although one might expect the greater time spent (and therefore greater sampling done) by high NFCs to lead to greater accuracy, it is clear from this analysis that high NFCs in this sample gained no significant benefit in accuracy.

2.3. Discussion

The clear and unequivocal conclusion that can be drawn from the analyses just discussed is that NFC is indeed related to the amount of time spent gathering information; in particular, as predicted, high NFCs[5] tend to spend longer seeking information than low NFCs, when cognitive ability is controlled. A surprising concurrent finding was that NFC in this sample was not significantly related to accuracy. In fact, it trended in the opposite direction than expected. One would expect that the relationship between NFC and time taken might have led to higher accuracy as a result of larger samples gathered, but this causal structure was certainly not apparent in my sample. Thus high NFCs, at least in this sample, do not appear to be benefiting from their extra effort. If it is the case that NFC is truly unrelated to accuracy, then the extra time spent at this task would appear to be wasted (and the performance, therefore, non-normative). Another possibility is that, in general, people do achieve accuracy benefits, but these are slight, in accordance with the diminishing returns of sampling (Hertwig & Pleskac, 2008).

What reason might high NFCs have for taking longer to sample? I will consider several. First of all, given that high NFCs did not achieve higher accuracy

[5] Although I always, when possible, examine the relationship between NFC and time taken when cognitive ability is partialed out, for ease of discussion I will henceforth use the term "high NFCs" to refer to the behaviour of high NFC individuals as a result of this unique contribution of NFC to time taken.

than low NFCs, they may require this extra time to achieve similar accuracy to others (an unlikely possibility, given NFC's relationship to cognitive ability, but important to rule out). Second, as I outlined in my predictions, they may be prone to adopt different strategies; they may focus more on accuracy, attempt to improve at the task, search for patterns where none exist, or otherwise attempt to remember extraneous information about the task. Also, they may lack motivation for the task, becoming distracted due to boredom, or their motivation may be only to please the researcher, in which case they spend extra time because they believe it will make them appear in a more positive light. Finally, they may simply possess a rule of thumb that suggests to them that sampling more is always better.

2.3.1. Ability failure

One potential reason for high NFCs' apparent time-wasting is that they are simply in need of the extra time in order to reach a comfortable level of accuracy. One possible – though unlikely – reason for this is that they may be better suited to more complex problems than this one, due to greater experience and enjoyment with such. I.e., perhaps, due to lack of experience, they are unused to the simple task of remembering and comparing information, and it takes them more time than others to reach a similar level of accuracy. This possibility is unlikely because complex problems also typically involve simpler parts, including remembering and comparing information.

However, regardless of potential reasons for this kind of failure, it is important to examine it as a possibility, especially given that an *ability* failure should be very different from a *strategy* or *motivation* failure like those mentioned above (e.g., a strategic focus on accuracy, or boredom). Such an ability failure is not irrational precisely because it is not a failure of strategy, as normativity is defined as taking the

best possible strategy to reach one's goals (or, in this case, the experimenter's goals). High NFC participants may have identified and pursued a normative strategy but simply possessed cognitive limitations preventing them from achieving as much success as others. The hypothesis that high NFCs suffer from an ability failure is thus the focus of Experiment II, wherein I asked participants to sample faster in a second, otherwise identical block. If their failure is one of ability, the high NFCs should perform worse when their sampling time is curtailed, in comparison to a control condition.

2.3.2. Strategy differences: Accuracy focus

The clearest and simplest reason people spend longer sampling, in general, is to trade speed for accuracy. Considering that high NFCs in this sample reported a focus on accuracy in general, there is a strong possibility that this was the goal of their extended time sampling, even if it was not (apparently) successful. Such a focus on accuracy is in accordance with their reported enjoyment of spending cognitive effort: Maybe they enjoy spending this effort because, in general, they are pleased by attaining accurate or correct results. Alternatively, perhaps they aim for accuracy because they are enjoying the task and truly engaging with it; they know they are enjoying spending time so they use the time to try to be more accurate.

If successful, the trading of speed for accuracy can be considered normative according to the participant's individual goals, assuming he values accuracy enough. Notably, though, in Chapter 1 I highlighted several reasons to be skeptical about an overwhelming focus on accuracy, in particular the high premium paid for accuracy thanks to the diminishing returns. Thus this strategy may be normative for a given participant, but it may also be that a focus on accuracy tends to get the participant "carried away" – and he overspends time without realizing it. In addition, recall my

discussion in Chapter 1 of participant-defined vs. experimenter-defined normativity. Given that there was no positive relationship between accuracy and NFC for this task, it is apparent that the high NFCs achieved a worse accuracy/speed ratio than low NFCs, who spent less time but were equally accurate – thus a focus on accuracy in this task leads to non-normativity, in the experimenter-defined sense. Since NFC is correlated with self-reported accuracy focus, I argue that accuracy focus is the strongest possible reason for the NFC-time relationship thus far, and it therefore should certainly be examined further.

Although my mediation analysis did not reveal accuracy focus as a significant mediator between NFC and time taken, there could have been a number of reasons for this result. First, my self-report measure reflects only what the participants report and is not validated, and second, a null result may always be the result of a Type II error. One way to overcome these problems is to examine accuracy focus behaviourally. A simple way of doing so is to determine participants' thresholds; e.g., perhaps one participant wishes to attain 80% accuracy and samples until the evidence is strong enough to indicate this, while another wishes 90% accuracy and therefore spends longer sampling in order to achieve this higher threshold of evidence strength. I will therefore look at participants' thresholds in Experiment III to determine if indeed high NFCs have higher thresholds than low NFCs. I also designed Experiment V to test whether high NFCs behave similarly to participants instructed to focus on accuracy (and whether low NFCs behave similarly to those instructed to focus on efficiency).

Given that most decision models postulate that confidence (in one's accuracy) rises according to evidence strength (e.g., Busemeyer & Townsend, 1993; Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2009; Ratcliff & Starns, 2009), a complementary question is whether the confidence of high NFCs grows in a different

way to that of low NFCs, perhaps according to a shallower curve. That is, if one asks participants to track their confidence until a threshold is reached (i.e. a choice made), high NFCs may show a shallower (and longer lasting) confidence curve than low NFCs due to a cautious nature preventing them from being more confident in the evidence they witness as they go along (although the two groups' confidence after the choice is made may be indistinguishable). I designed Experiment IV to examine participants' confidence growth. This design may also reveal any abnormalities in the way that various individuals, in particular high NFCs, become confident in response to data.

2.3.3. Strategy differences: Improvement

As I mentioned, high NFCs in this sample were equally (perhaps even less) accurate than their low NFC peers, when cognitive ability was accounted for, and it is therefore entirely possible that they do not achieve greater accuracy in general. An important idea to consider, then, is that the high NFCs may be dividing their attention. I.e., part of their attention goes to remembering evidence seen, and part toward fulfilling some other goal. One possibility is that they are actively seeking strategies that will help them improve on this task. This meta-strategy is of great interest because it is what I would call the wisest strategy: It has the potential to optimise over the long term of the problem. Granted, the participant is not likely to be faced with this particular problem for very long, so within the experiment it is not necessarily a normative strategy; it depends on one's judgment of how long the experiment will last, and therefore how much time could be saved by learning to improve. However, importantly, an attempt to improve, which only fails due to a judgment error, could still be considered a normative strategy (as the strategy itself has the potential to optimise, even if one's ability does not allow it). Further, a tendency to search for the

best long-term strategy could certainly be considered a mark of wisdom in everyday life. Although I did not find any relationship between NFC and improvement in this experiment, I continue to examine it in subsequent experiments.

2.3.4. Strategy differences: Pattern-searching

There is a second possible way that high NFCs may be dividing their attention, one that is further from normativity. It is possible that they are distracted by attempting to search for patterns, applying an erroneous problem-solving rule to the randomly generated data they are receiving. It has been argued in the past that many of the biases humans display are due to such misapplication of learned rules (e.g., Arkes & Ayton, 1999), especially given that nonhuman animals often do not display human biases – for example, pigeons perform more optimally than humans on matching-to-sample tasks (Goodie & Fantino, 1995; Hartl & Fantino, 1996), base-rate problems (Goodie & Fantino, 1995), sunk cost problems (Arkes & Ayton, 1999), and the Monty Hall dilemma (Herbranson & Schroeder, 2010), and most nonhuman animals outperform humans on probability guessing tasks (Hinson & Staddon, 1983).

To give a detailed example, consider probability guessing, a similar problem to the current scenario. In this type of problem, participants are presented with stimuli that appear with a certain frequency (e.g., red appears 75% of the time, green appears 25% of the time), and they must guess which one will appear on a subsequent trial. Most humans exhibit matching behaviour; i.e., they will be sensitive to the appropriate frequencies but attempt to predict new stimuli according to a pattern, generally guessing red 75% of the time and green 25% of the time (Yellot, 1969). This strategy is, notably, not the optimal one, which is to choose red on every trial (for 75% accuracy). People readily report that pattern discovery is their goal, as when Yellot changed the final trials of such an experiment so that the participant's guess

was always correct, many participants' interpretations were that they had finally found the correct pattern. Given the preference of high NFCs for complex problems and a life filled with puzzles, it is thus entirely possible that high NFCs spend extra time searching for patterns in the data they receive, perhaps within a given trial or between trials (to see, e.g., if Candidate A is generally better than Candidate B). The reason such a pattern-searching strategy is an especially strong possibility in the case of high NFCs is that problem-solving behaviour is dependent on one's past problem-solving experience. High NFCs' enjoyment of complex problem solving implies experience with such, as an individual is likely to seek out what she enjoys, and it's therefore even more likely that they have developed a pattern-searching habit than low NFCs.

And yet, people who have become used to searching for patterns in some problems may erroneously apply this rule on other problems (see Fantino, Kanevsky, & Charlton, 2005, for an interesting example of pigeons developing non-normative behaviour after being trained in a habit only beneficial to a different task). Such pattern searching is, naturally, fruitless on this task, as data were generated randomly, and participants had no reason to believe otherwise. And the extra cognitive resources required to do so (e.g., memory expended trying to remember the pattern of A and B candidates chosen in the past) could easily have made determining the better candidate more difficult, as there would be less room in the memory store for new evidence (faces). Even if high NFCs are not explicitly searching for patterns, any attempt to store extra information, which they might desire to do, could cause such difficulty. I designed Experiment VI to test participants in a "sampling" scenario in which they are less likely to search for patterns (thinking of numerical answers to trivia questions, e.g., "What percentage of countries are African?"). Because any

sampling done is of items in memory, there should be no tendency to seek patterns in that information, nor is it possible to gather any new information to remember, nor is it likely one could focus energy into improving at the task. Thus a propensity to think for longer, on that task, should instead be the result of a focus on accuracy or a general heuristic to spend longer.

2.3.5. Motivation differences: Boredom

Another non-normative possibility is that high NFCs, with their given enjoyment of complexity, are simply bored with this task and using extra time to think of *anything* else – i.e., they are not focusing cognitive effort on pattern-searching or task-related activities, but are instead putting that effort toward other, unrelated thoughts. If they are still attempting a reasonable level of accuracy, this may result in more time spent sampling, as again their distraction should hinder their ability to assess the evidence. This behaviour is clearly non-normative, as it does not optimise for the goals set out by the experimenter: to be both accurate and efficient. Some time is "wasted" in thinking of things not related to these goals. Such behaviour should also be eliminated in the short-term sampling problem presented in Experiment VI, as the fast pace of problems should make it difficult to maintain thoughts of other things. Thus, again, a tendency to think for longer on the short-term sampling problems should be due to a focus on accuracy or a time-spending heuristic.

2.3.6. Motivation differences: Social desirability

Another possibility that is apparently non-normative is that high NFCs spend longer because they simply believe that it is expected of them, and that this behaviour will therefore create a more positive impression on the researcher. In other words, it is entirely possible that, in a lab setting, the interpretation of the participant is that the

researcher desires her either to be as accurate as possible, or at least to expend a certain amount of effort (i.e., time).

Although there was no significant relationship between NFC and SDR in this sample, they are typically significantly (if weakly) correlated in the literature (Cacioppo et al., 1996), and the low reliability of the SDR measure indicates that it may not have been accurately representing the true underlying variable. It is therefore still possible that high NFCs may attempt to spend more effort in order to create a positive impression.

Perhaps they might do so, as I have already discussed, in order to focus on accuracy (because they believe accuracy is desired) – this behaviour is not normative for the same reason discussed above: because they fail to achieve a good accuracy/speed ratio. Their assignment of utility, in this case, is clearly not ideal, and this assignment is precisely what determines their strategy. Alternatively, perhaps participants interpret only a desire for them to spend effort/time and not necessarily to achieve greater accuracy; in this case, they may simply spend this extra time without using it to think about the task – instead using it to think about other things, while giving the *appearance* of expending effort. This behaviour is certainly non-normative because, again, it does not optimise for the experimenter's set goals of accuracy and efficiency, as time is spent on thoughts unrelated to the task. Experiment VI's alternate sampling task should again be well suited to testing this hypothesis, as participants are unlikely to spend more time on very short trivia questions in order to appear to be expending effort. I will also examine any further case of an NFC-time relationship to determine if it is mediated by SDR.

2.3.7. A heuristic for spending time

The final possibility I will consider as a potential reason for the NFC-time relationship is that high NFCs have a simple heuristic to spend time on problems. Such a heuristic states that *the entire goal is to spend time.* In other words, spending time is not undertaken because of, e.g., a focus on accuracy, or in order to search for patterns, though these may arise as side-effects (as something to fill the time already committed). This possibility is also functionally indistinguishable from an inability to *disengage* from the task. All in all, the participant may be spurred on by a continued feeling to spend more time on the problem, making it difficult, at any point, to stop. This hypothesis is in line with the suggestion by Tanaka, Panter, and Winborne (1988) that there are several sub-scales of NFC, one of which they labeled Cognitive Persistence.

Naturally, the spending of time for no reason other than to spend time is, for this problem, not a normative strategy, even though it may result in side-effects that could be useful on other problems. This particular possibility is the most difficult to test, as it is difficult to distinguish from its potential side-effects. However, an NFC-time relationship discovered in the small-scale sampling task of Experiment VI would, as argued above, most likely be the result of either accuracy focus or just such a heuristic.

2.3.8. Summary

In sum, I have found a clear relationship between NFC and time taken to gather information, when cognitive ability is controlled, such that high NFCs take longer to gather evidence without any apparent benefit. The importance of this discovery by itself is apparent: It suggests that the NFC scale can be used to predict

decision making behaviour, and, although I cannot draw firm causal conclusions about the relationship, it also suggests some caution in pursuing effortful thinking in all situations. Although NFC uniquely predicts normative performance on several difficult tasks on which participants typically show bias (e.g., Kokis et al., 2002), in this case it appears NFC predicts poorer overall performance (equal or lesser accuracy but more time spent).

I have also enumerated several possible reasons for this relationship, including one that is ability-based (Ability Failure), three that are strategy-based (Accuracy Focus, Improvement, Pattern-searching), two that are motivation-based (Boredom, Social Desirability), and one that is heuristic-based (Time Heuristic). Importantly, I have argued that only two of these possibilities, Ability Failure and Improvement, would be consistent with normative behaviour on this task. Determining which reason, or combination of reasons, is the most likely cause of the NFC-time relationship is thus crucial both for understanding deliberative processes in general, and where and when they lead to normativity. The determination of these reasons is the focus of the rest of this thesis.

3. Experiment II: Becoming more efficient

One potential explanation I mentioned for the results of Experiment I was that individuals high in NFC were at an ability disadvantage in this task. In other words, they may need more time than others to reach an equivalent (or lower) level of accuracy; perhaps they are attempting to reach the same level of accuracy as others but are simply constrained to take a longer path to get to it. Such a finding would indicate that they are not exhibiting time-wasting behaviour, as might be the case if they were searching for nonexistent patterns, pushing for only negligible improvements in accuracy, or lengthening their search due to boredom, a heuristic, or to please the experimenter. It is important to distinguish between these two types of possibilities, because, in the former case, it's quite possible that all are behaving according to the same strategy: They have chosen a certain level of accuracy that is desirable and sample until they reach it. Meanwhile, in the latter cases, the high NFCs are doing something unusual. In these cases, their approach to the task may differ substantially, as would be the case with a pointless search for extraneous information, or reflect a difference in parameters, as in the case where the level of accuracy they desire is simply higher.

Experiment II was designed to test for a difference in ability. It replicates Experiment I with one major change: There were two blocks, and in the second block some participants were told to gather information faster (Speed condition), while the rest continued sampling as before (Control condition). If high NFC individuals are simply poor at this task, then their accuracy when switching to sampling faster should plummet compared to their accuracy at the start; they should also have lower accuracy scores, in the second block, than high NFCs who receive no special instructions. If, however, they are able to perform well when sampling faster, then there is likely some

other cause for their extended sampling and poor performance in Experiment I, perhaps a focus on accuracy or a propensity to search for patterns in the evidence.

In sum, the purpose of Experiment II was to compare both between and within participants for differences in accuracy – i.e., to determine both a) if high NFC participants in the Speed condition are less accurate than their peers in the Control group and b) if they are less accurate in the speeded Block 2 than they previously were in Block 1. In either of these cases, I could conclude that high NFCs find this task difficult and are sampling more to compensate. Low NFCs, meanwhile, should show no change in accuracy across either conditions or blocks.

3.1. Method

3.1.1. Participants

30 adults (25 female and 5 male) were recruited from the Cardiff University Human Participants Panel and received course credit for their time. One participant did not complete the questionnaires and was removed from analysis. Of those who were later included in the analysis, the median age was 19.

3.1.2. Procedure and choice task

Participants underwent the same procedure and completed the same choice task as in Experiment I, with a few exceptions. First, participants completed a total of 40 trials instead of 20 (the second 20 being identical in nature to the first). Second, they were randomly assigned to one of two conditions. In each condition, after 20 trials, the participant received additional instructions. In the *Speed* condition, the instructions were to focus on "efficiency over accuracy"; this was clarified by asking them to "still try to make the right choice, but worry more about moving through

candidates quickly." In the *Control* condition, participants were simply asked to take a short break before continuing.

A third difference was that participants completed a second task after the choice task; this task involved a separate experimental design and will therefore be discussed as a separate experiment (Experiment III).

Finally, in order to better ensure that the participants' answers to the questionnaires were not influenced by the choice task itself, I asked participants to return between 2 and 14 days after the choice task to complete the NFC and SDR questionnaires (although they completed the Raven matrices on the same day).

3.1.3. Materials

The Cognitive Ability, NFC, and SDR materials in this experiment were identical to those in Experiment I. Mean scores were –1.43 (*SD* = 6.82) for the Raven matrices, 87.07 (*SD* = 11.13) for the NFC measure and 16.07 (*SD* = 5.84) for the SDR measure; these means did not differ significantly from those in Experiment I (all *z*s < 1.93 in magnitude). Cronbach's α for these measures were .71, .89, and .74 respectively.

3.2. Results

I removed one participant from analysis for appearing as an outlier in several correlations and an initial regression analysis of NFC and Cognitive Ability onto Time Taken, designed to identify such outliers. The final number of included participants was thus 28. As in Experiment I, I considered single-judgment trials to be errors, which I discarded from analysis. The median number of such errors per participant was 2, and the greatest number of errors made by a single participant was 7 (in all 40 trials).

All dependent measures (Time Taken, Accuracy, Median Confidence, Improvement, AE Focus – Task, and AE Focus – General) were measured in the same way as in Experiment I, and, for Block 1 (which replicated Experiment I), only AE Focus (Task) and Time Taken had mean scores significantly different from those in Experiment I (for the others, all zs < 1.93 in magnitude). For AE Focus (Task), participants' average focus in this experiment was 36.36 ($SD = 12.11$), which, unlike in Experiment I ($M = 50.79$, $SD = 20.96$), was significantly different from the mid-point of 50, $t(27) = -5.96$, $p < .001$, indicating a focus on accuracy despite the equal emphasis on accuracy and efficiency in the instructions; this represented a significant deviation from Experiment I's scores, $U = 221.50$, $z = -2.80$, $p = .005$. Perhaps participants in this study, drawn from an undergraduate credit panel instead of a paid participant panel, considered the setting of the study (the building of their School) to be more evaluative and therefore interpreted instructions in a way that emphasized accuracy. Additionally, participants in this experiment spent significantly less time in Block 1 ($M = 2.72$, $SD = 0.12$; scores logged) than did participants in Experiment I ($M = 2.79$, $SD = 0.13$), $U = 259.0$, $z = -2.18$, $p = .03$. Again, this difference may have been due to the different samples; paid participants may be more motivated to spend time in order to earn their payment than participants who must fulfill a participation requirement for their School. Given that I would still predict individual differences in time spent even within these two potentially distinct populations, I proceeded with the analysis. Descriptive statistics for all dependent measures are reported in Table 3.1.

Because I intended to include Cognitive Ability as a covariate in the main analysis, I conducted a univariate ANOVA with Cognitive Ability as the dependent measure, including condition and NFC (measured) as the independent variables. This analysis revealed no main effects or interactions (all Fs < 0.40), indicating that

Cognitive Ability is appropriate to include as a covariate, as it does not appear to be significantly related to NFC in this sample and therefore should exhibit a unique effect on any outcome variable. Correlations between Cognitive Ability and primary variables in this study are displayed in Table 3.2. It is of some note that Cognitive Ability was not related to these variables in the same way as in Experiment I; in particular, it was not correlated with NFC, $r = -.11$, $p = .60$, or Time Taken in Block 1, $r = -.09$, $p = .64$, or Block 2, for either condition, $r = -.03$, $p = .93$, for the Control condition and $r = -.17$, $p = .58$, for the Speed condition. Neither was it correlated significantly with Improvement in either Block 1, $r = -.32$, $p = .10$, or Block 2 Control, $r = -.08$, $p = .81$, though the correlation was marginal in Block 2 for the Speed condition, $r = .45$, $p = .09$. Cognitive Ability was only significantly correlated with AE Focus (General), $r = -.64$, $p < .001$, indicating that participants high in Cognitive Ability reported preferring to focus on accuracy in general. Further, Cognitive Ability did not contribute significantly to any of the following analyses as a covariate (all Fs < 3.97). These results will be discussed in more detail later.

Table 3.1. Experiment II: Descriptive statistics for dependent measures, by condition.

	Block 1		*Block 2 - Control*		*Block 2 - Speed*	
	Mean	SD	Mean	SD	Mean	SD
Time Taken (log)	2.72	0.12	2.68	0.16	2.55	0.09
Accuracy	9.57	3.35	10.38	2.06	10.20	4.38
Median Confidence	11.50	2.65	11.31	3.10	8.70	3.64
Improvement	-.01	.26	-.07	.24	-.04	.33
AE Focus - Task	36.36	12.11	56.85	15.83	74.67	12.82
AE Focus - General	39.39	17.91	-	-	-	-

Note. AE Focus refers to the participant's focus on accuracy or efficiency; a high score indicates a focus on efficiency, while a low score indicates a focus on accuracy. AE Focus (General) measures this focus in general and thus only one score was recorded per participant; this is listed under Block 1.

Table 3.2
Experiment II: Intercorrelations among primary variables.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Ability Measures																	
1. Cognitive Ability	–																
2. Need for Cognition	-.11	–															
Choice Task - Block 1																	
3. Time Taken (log)	-.09	.51 **	–														
4. Accuracy	-.30	.01	.19	–													
5. Median Confidence	.15	-.22	-.15	-.10	–												
6. Improvement	-.32 §	.36 §	-.06	-.16	-.18	–											
7. AE Focus - Task	-.23	-.35 §	-.47 *	-.13	-.29	.27	–										
Choice Task - Block 2 (Control)																	
8. Time Taken (log)	-.03	.73 **	.95 **	.26	-.34	-.03	.18	–									
9. Accuracy	.38	.57 §	.43	.03	.18	.16	-.13	.45	–								
10. Median Confidence	.29	-.14	.02	-.36	.66 *	-.13	-.14	-.13	.04	–							
11. Improvement	-.08	.38	.19	-.17	.06	.40	.59 *	.30	.55 §	-.05	–						
12. AE Focus - Task	.53 §	.07	.23	.28	.12	-.70 **	-.44	.18	.01	.13	-.44	–					
Choice Task - Block 2 (Speed)																	
13. Time Taken (log)	-.17	.43	.81 **	.24	-.61 *	-.30	-.18	–	–	–	–	–	–				
14. Accuracy	.21	.24	-.11	-.52 §	.22	.00	-.68 *	–	–	–	–	–	.04	–			
15. Median Confidence	-.08	-.02	-.58 *	-.21	.67 **	-.28	.01	–	–	–	–	–	-.41	.09	–		
16. Improvement	.45 §	.51 §	-.21	-.12	.52 *	-.25	-.26	–	–	–	–	–	-.02	.01	.27	–	
17. AE Focus - Task	.02	-.41	.09	.15	.34	-.71 **	-.05	–	–	–	–	–	-.18	-.48 §	.20	-.10	–
AE Focus																	
18. AE Focus - General	-.64 **	-.40 *	-.18	-.02	-.24	.31	.36 §	-.13	-.14	-.46	.39	-.51 §	-.29	-.46	.13	-.26	.22

Note. Outliers were assessed and removed in pairwise correlations by the Jackknifed Mahalanobis distance method (Mahalanobis, 1936; also see section 1.7.1). Time Taken refers to the total amount of time taken to complete all 20 trials in a block. AE Focus refers to the participant's focus on either accuracy or efficiency; a low AE Focus score indicates a focus on accuracy and a high score indicates a focus on efficiency. I do not report correlations between the two conditions, as these were between-subjects.

§ $p < .10$.
* $p < .05$.
** $p < .01$.

Correlations among the other primary variables are displayed in Table 3.2. Many of the significant relationships listed are better described by the ANCOVA analyses I do in the next section; however, there are some notable trends that I will describe here. In particular, I found that NFC is significantly correlated with AE Focus (General), $r = -.40$, $p = .045$, such that those high in NFC were more likely to report a general focus on accuracy, as in Experiment I. Also similarly to Experiment I, the more time participants took in Block 1, the more likely they were to report a focus on accuracy within that block, $r = -.47$, $p = .02$, although this relationship was not apparent in Block 2, for either condition. Unlike previously, there was no relationship between time spent and median confidence in Block 1, $r = -.15$, $p = .46$.

Some new relationships also emerged. First, more time taken during Block 1 was associated with lower confidence during Block 2 for the Speed condition, $r = -.58$, $p = .02$. It is possible, then, that those who spent more time in Block 1 were more likely to cut out a lot of time in Block 2, and thus feel less confident. Conversely, the more confident one was during Block 1, the less time one took when asked to sample faster in the Speed condition of Block 2, $r = -.61$, $p = .02$, possibly because more confident individuals think they are able to do better with less evidence. Further, greater confidence in Block 1 was also associated with more improvement in Block 2's Speed condition, $r = .52$, $p = .046$; perhaps such confident participants overshot their ability and then improved in Block 2. There was also a marginal negative correlation between accuracy in Block 1 and accuracy in Block 2 for the Speed condition, $r = -.52$, $p = .055$, meaning that the more accurate one was to begin with, the lower one's accuracy when forced to truncate one's sampling.

Finally, there were some relationships that were more puzzling. For example, the more one focused on efficiency during Block 1, the more improvement one

showed during Block 2, for the Control condition, $r = .59$, $p = .04$, and the less accuracy one showed during Block 2, for the Speed condition, $r = -.68$, $p = .01$. This may be because, in the former case, an early focus on efficiency left room for more improvement later. In the latter case, I simply cannot speculate. Also, the more one improved during Block 1, the more one was likely to focus on accuracy during Block 2, for both the Control condition, $r = -.70$, $p = .008$, and the Speed condition, $r = -.71$, $p = .007$. These relationships may be evidence that improvement brings positive self-evaluation, which makes one enjoy accuracy more (and therefore focus on it).

3.2.1. Main analyses

I first examined the effect of NFC, condition, block, and evidence universe on the amount of time taken to complete a block using a 2 (Condition: Control or Speed) x 2 (Block: 1 or 2) x 2 (Universe: Weak or Strong) x NFC (measured)[6] ANCOVA, with Cognitive Ability as a covariate. Figure 3.1 displays the expected pattern of results. First, the figure shows the expected pattern of difference between high NFCs and low NFCs; it is clear that those high in NFC take longer overall (regardless of condition), $F(1, 23) = 5.86$, $p = .02$, $\eta_p^2 = .203$. The analysis also revealed a significant block by condition interaction, $F(1, 23) = 23.55$, $p < .001$, $\eta_p^2 = .506$. It is apparent that participants in the Control condition spent roughly the same amount of time in both blocks, while those in the Speed condition dramatically reduced the amount of time spent in Block 2. Simple effects analyses suggested both that a) participants instructed to sample faster in Block 2 indeed did so (for those in the Speed condition, $M = 2.42$, $SE = 0.03$, in Block 1, vs. $M = 2.25$, $SE = 0.03$, in Block

[6] Again, all analyses in this section were also performed using a median split of NFC (*Median* = 84.0), with a similar pattern of results. Only in one case was there a slight difference: For the current analysis, there was only a marginal main effect of NFC, $p = .063$.

2), $F(1, 23) = 84.37$, $p < .001$, $\eta_p^2 = .786$, and b) participants in the Speed condition moved through Block 2 faster ($M = 2.25$, $SE = 0.03$) than those in the Control condition ($M = 2.42$, $SE = 0.03$), $F(1, 23) = 7.03$, $p = .01$, $\eta_p^2 = .234$. I conducted further analyses to be certain that this pattern of results was similar for both high and low NFCs. Examination of Figure 3.1 shows that this appears to be the case; the flat trend in the Control condition and the declining trend in the Speed condition are similar for both high and low NFCs. Accordingly, the block x condition x NFC interaction was not significant. Thus the manipulation of the Speed instructions was clearly successful, and for both high and low NFCs.

Further, I found the predicted main effect of universe, $F(1, 23) = 7.54$, $p = .01$, $\eta_p^2 = .247$, such that participants spent more time in the weak universe trials ($M = 2.38$, $SE = 0.02$) than the strong universe trials ($M = 2.35$, $SE = 0.02$). However, the relationship involving evidence universe in this experiment is best described by a three-way interaction: The interaction between block, universe, and NFC was significant, $F(1, 23) = 5.15$, $p = .03$, $\eta_p^2 = .183$. Figure 3.1 shows that only the high NFCs spent significantly longer in the weak evidence trials, and simple effects analyses confirm this result. They show a significant effect of universe within Block 1 for the high NFCs, $F(1, 23) = 16.83$, $p < .001$, $\eta_p^2 = .423$, and a marginal effect of universe within Block 2 for the high NFCs, $F(1, 23) = 3.95$, $p = .06$, with no significant effects of universe within either block for low NFCs (both Fs < 1.53). High NFCs thus spent significantly longer, at least in Block 1, on weak evidence trials (Block 1: $M = 2.46$, $SE = 0.03$, Block 2: $M = 2.35$, $SE = 0.03$) than on strong evidence trials (Block 1: $M = 2.41$, $SE = 0.02$, Block 2: $M = 2.33$, $SE = 0.02$); meanwhile, low NFCs spent about the same time on all trials (Block 1: $M = 2.38$, $SE = 0.02$, Block 2: $M = 2.27$, $SE = 0.02$). Overall, this interaction suggests that high NFCs differentiate

between evidence universes, spending more time when evidence is weak in order to gather more.

In order to examine whether accuracy would drop for those participants moving more quickly through the trials, I performed a similar ANCOVA analysis, except removing evidence strength as a factor[7], and with Accuracy as the dependent variable. Examination of Figure 3.2 shows that, clearly, accuracy remained the same or even improved in all cases, regardless of NFC or condition; i.e., there were no significant main effects or interactions (all Fs < 1.72). It is especially important to note that, in the Speed condition, both those high in NFC and low in NFC maintained or even improved their accuracy in Block 2; performance was also extremely similar in the Control condition. There is thus no support from these data for the hypothesis that NFC is related to an ability deficit: Even when sampling much faster than they had previously (and compared to a control group), all participants in this sample were able to maintain a high level of accuracy.

I then conducted identical ANCOVA analyses on the three other dependent measures: Median Confidence, Improvement, and AE Focus (Task). First, the ANCOVA on Median Confidence (see Figure 3.3) revealed only a main effect of block, $F(1, 23) = 12.21$, $p = .002$, $\eta_p^2 = .347$, such that participants in both conditions had attenuated confidence in Block 2 ($M = 9.91$, $SE = 0.68$) compared with Block 1 ($M = 11.50$, $SE = 0.50$), and a main effect of condition, $F(1, 23) = 5.36$, $p = .03$, $\eta_p^2 = .189$, such that those in the Speed condition reported lower confidence ($M = 9.53$, $SE = 0.74$) than those in the Control condition ($M = 12.05$, $SE = 0.79$). There were no other significant effects (all Fs < 1.10).

---

[7] Note that since including evidence strength increases the complexity of the analysis, and since I had no further specific and pertinent hypotheses involving it, I chose to omit it from the remainder of the analyses.

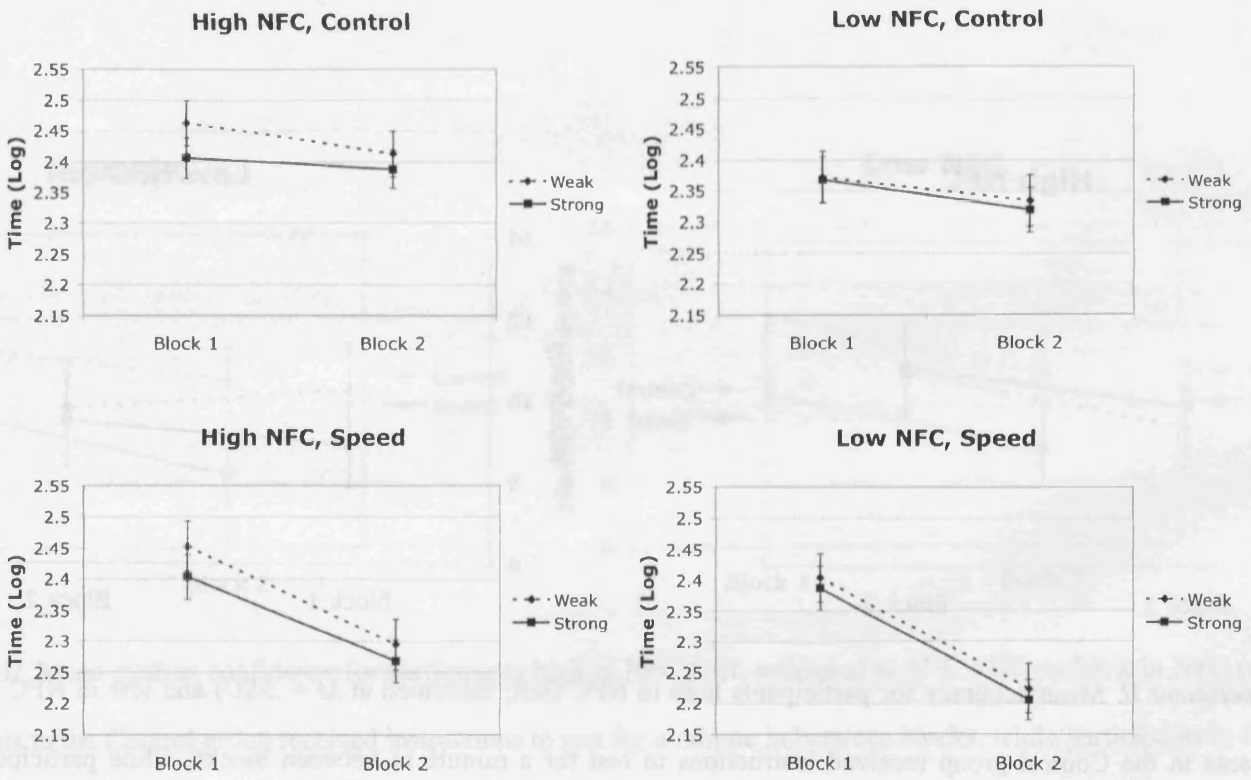


Figure 3.1, Experiment II. Mean log of time taken to complete the blocks (10 trials each for strong and weak universes) for participants high in NFC (left; estimated at $M + .5SD$) and low in NFC (right; estimated at $M - .5SD$). Participants in the Control group were instructed to rest for a minute in between blocks; participants in the Speed group were instructed to sample faster in the second block. Bars are 1 standard error.

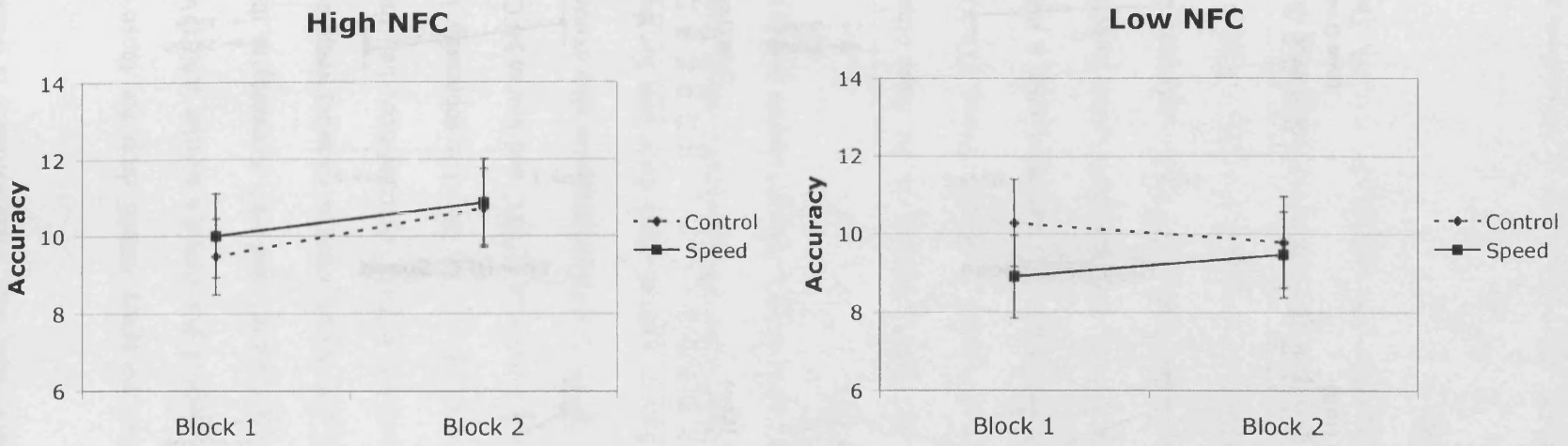


Figure 3.2, Experiment II. Mean accuracy for participants high in NFC (left; estimated at $M + .5SD$) and low in NFC (right; estimated at $M - .5SD$). Participants in the Control group received instructions to rest for a minute in between blocks, while participants in the Speed group received instructions to sample faster in the second block. Bars are 1 standard error.

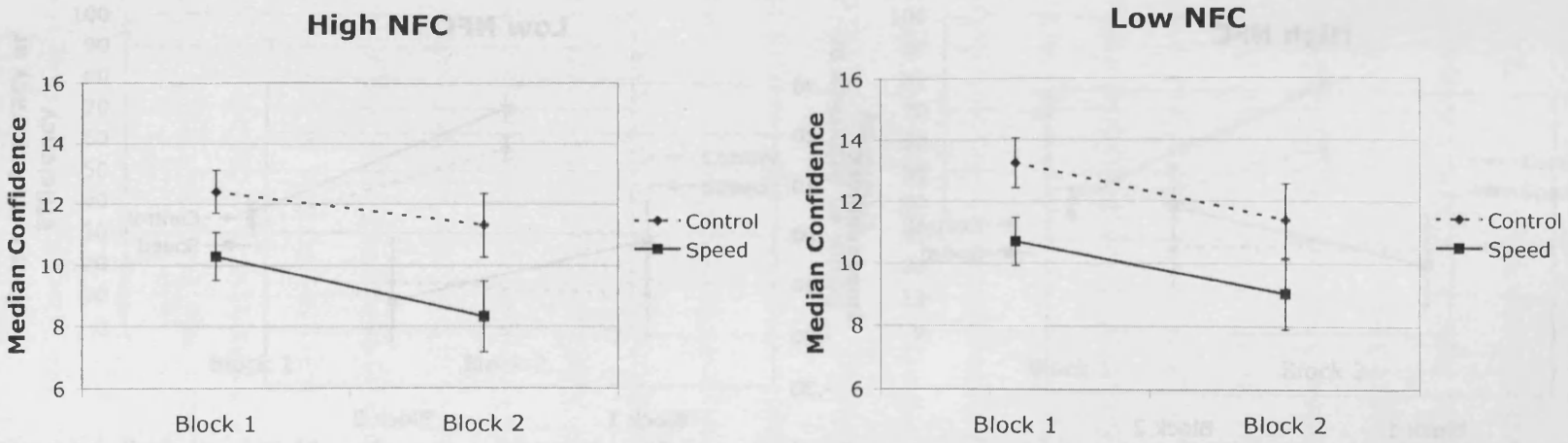


Figure 3.3, Experiment II. Mean median confidence for participants high in NFC (left; estimated at $M + .5SD$) and low in NFC (right; estimated at $M - .5SD$). Participants in the Control group received instructions to rest for a minute in between blocks, while participants in the Speed group received instructions to sample faster in the second block. Bars are 1 standard error.

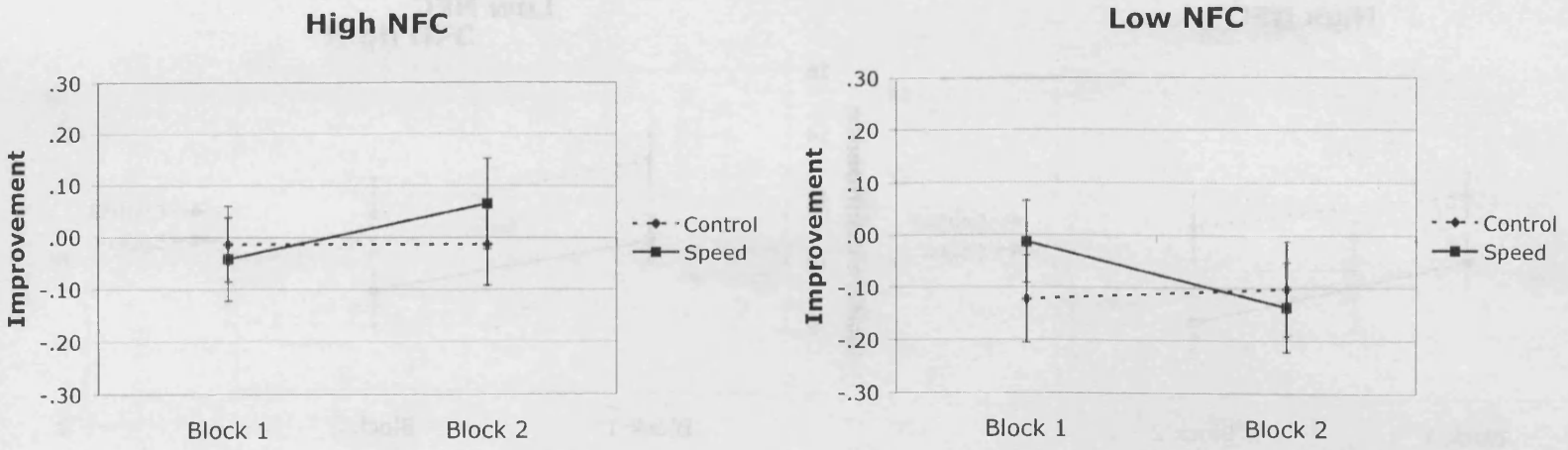


Figure 3.4, Experiment II. Mean improvement score for participants high in NFC (left; estimated at $M + .5SD$) and low in NFC (right; estimated at $M - .5SD$). Zero scores indicate no improvement; nonzero scores indicate improvement (positive scores) or decrement (negative scores) in performance. Participants in the Control group received instructions to rest for a minute in between blocks, while participants in the Speed group received instructions to sample faster in the second block. Bars are 1 standard error.

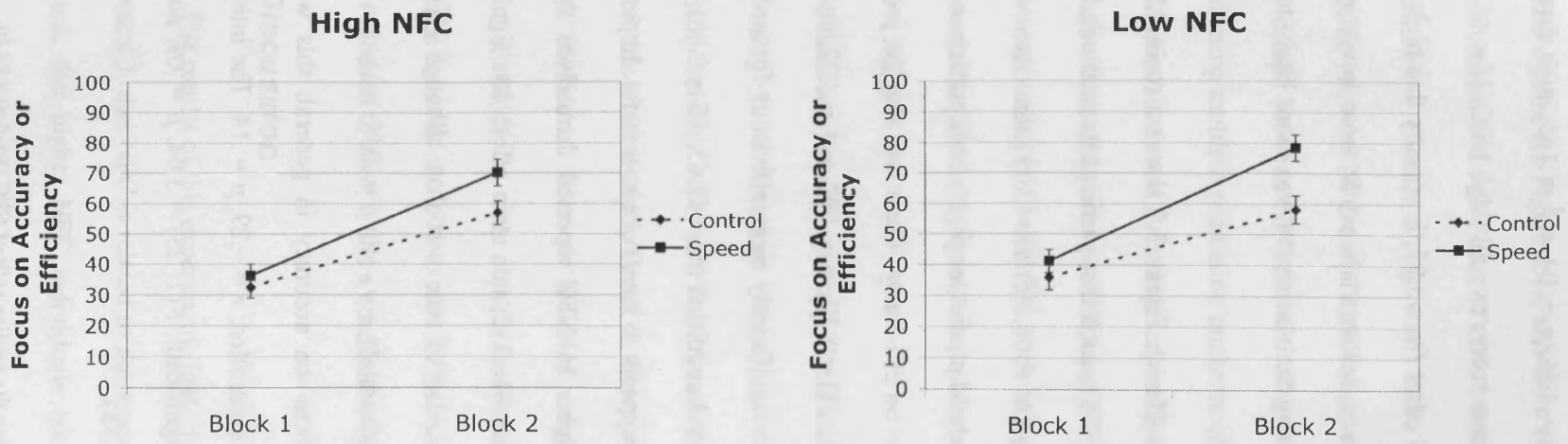


Figure 3.5, Experiment II. Mean focus on accuracy vs. efficiency during the task for participants high in NFC (left; estimated at $M + .5SD$) and low in NFC (right; estimated at $M - .5SD$). High scores indicate a focus on efficiency; low scores indicate a focus on accuracy; 50 is the mid-point and indicates equal focus. Participants in the Control group received instructions to rest for a minute in between blocks, while participants in the Speed group received instructions to sample faster in the second block. Bars are 1 standard error.

Second, the ANCOVA on Improvement (see Figure 3.4) resulted only in a significant main effect of NFC, $F(1, 23) = 4.56$, $p = .04$, $\eta_p^2 = .165$, such that those higher in NFC achieved higher improvement scores overall (high NFC: $M = .05$, $SE = .05$; low NFC: $M = -.14$, $SE = .06$)[8]; all other Fs < 3.97. It appears that high NFCs were more likely to maintain their accuracy across trials, while those lower in NFC achieved scores slightly below 0, indicating their accuracy was more likely to have attenuated slightly.

Finally, in the case of AE Focus (Task), Figure 3.5 shows a main effect of block, $F(1, 23) = 73.24$, $p < .001$, $\eta_p^2 = .761$, such that participants in Block 2 were significantly more focused on efficiency ($M = 66.39$, $SE = 3.11$) than they were in Block 1 ($M = 36.36$, $SE = 2.29$); as one would predict, as time passes participants tend to tire of the task and tend to focus more on efficiency. There was also the predicted significant main effect of condition, $F(1, 23) = 8.83$, $p = .007$, $\eta_p^2 = .277$, such that participants in the Speed condition were significantly more efficiency-focused ($M = 55.94$, $SE = 2.43$) than those in the Control condition ($M = 45.65$, $SE = 2.60$); again, the manipulation of the Speed condition appears to have been successful. Importantly, Figure 3.3 finally shows that those higher in NFC reported themselves as more focused on accuracy in both Block 1 and Block 2; this main effect was significant, $F(1, 23) = 4.58$, $p = .04$, $\eta_p^2 = .166$. Also of note was that, although NFC was significantly correlated with AE Focus (General), $r = -.40$, $p = .045$, such that higher NFC participants reported a higher focus on accuracy in general, this was not significant when Cognitive Ability was controlled, $r = -.29$, $p = .14$. The interaction between block and condition was not significant, $F(1, 23) = 3.04$, $p = .09$, nor were there other significant effects (all Fs < 1.58).

---

[8] These means are based on the analysis done using the median split of NFC (*Median* = 84.0).

Because there was a significant main effect of NFC on Time Taken (i.e., a significant *c* path; see Figure 2.1), and there were also significant main effects of NFC on Improvement and AE Focus (Task), I set out to determine whether Improvement and AE Focus (Task), along with SDR, might be potential mediators of the relationship between NFC and Time Taken. I began by searching for significant path *a* relationships between NFC and the other variables within each block, which I did by conducting multiple regression analyses of NFC onto each, with Cognitive Ability removed in all cases and Condition controlled for in Block 2 analyses. I did not find a significant relationship between NFC and AE Focus (Task), for either block (both *t*s < 1.67 in magnitude), but I did find such a significant relationship between NFC and Improvement in Block 2, regression β-weight = .41, $p = .04$, and between NFC and SDR, $\beta = .45$, $p = .02$.

I then conducted bootstrapped mediation analyses (Preacher & Hayes, 2004, 2008), using 5000 bootstrap resamples, to determine if a) SDR mediated the relationship between NFC and Time Taken in Block 1 and b) SDR or Improvement mediated this relationship in Block 2. There was no significant mediation in the latter case (all confidence intervals included zero), but SDR proved to be a significant mediator in Block 1, with a point estimate of .0017 and a 95% BCa bootstrap confidence interval of .0001 to .0055. This mediation was only partial, as the c' path was not reduced to zero; it had a coefficient of .0027.

3.3. Discussion

The results of this experiment, first of all, clearly replicate the main finding of Experiment I: that NFC is related to the amount of time spent on this choice-making task, such that higher NFC individuals take longer to sample information when

cognitive ability is controlled. Importantly, in the second block, I asked half of the participants to sample more efficiently – and not only were high NFC individuals clearly able to do so, they did so with no significant effect on their accuracy score (similarly to low NFCs). Thus it is not the case that the high NFCs in this sample had difficulty with this task, such that they were sampling to compensate: When they spent less time sampling, they were equally accurate. Since the results of the accuracy analysis are null results, however, it is not possible to argue conclusively that I would expect this pattern to generalize to the population. But these results serve as an existence proof that high NFCs certainly *need* not do poorly when sampling more efficiently. There is also clearly no support for the hypothesis that high NFCs have a deficit in ability.

It should be noted, of course, that there were several obvious differences between the two samples that I took, in this experiment and the former, which I have thus far tentatively attributed to differences in participants (paid in Experiment I, credit panel in Experiment II). Participants in the current experiment reported that they were more focused on accuracy during the task, took overall less time to complete the task in Block 1, and did not show the same pattern of relationships between cognitive ability and the dependent measures; most notably, there was no relationship between cognitive ability and time taken, for which there was a significant negative relationship in Experiment I in the regression with NFC. There was also significant mediation of SDR on the NFC-time relationship in this experiment, where there was not in Experiment I; this may be because students are more interested in pleasing their experimenter (or are generally more focused on appearing in a positive light to their superiors) than participants from a paid panel. Finally, in this sample I found that high NFCs were spending differentially more time

on weak universe trials, while low NFCs spent the same amount of time on the weak and strong evidence trials; previously there was no such difference between high and low NFCs. It is worth considering, then, that these populations, paid and participant panel, may be significantly different when performing on this task. Future research, outside the scope of this thesis, will be needed to determine if this is the case. However, one thing is clear: The relationship between NFC and time taken, with cognitive ability partialed out, remained strong in Experiment II, as in Experiment I. It is therefore likely that, whatever differences exist between the two samples, this phenomenon is a general one.

In my discussion of Experiment I, I mentioned several possible reasons for the relationship between NFC and time taken. Namely, I suggested that it could be that participants have a focus on accuracy, are thinking about other things (either unrelated, e.g. bored daydreaming, or related to the task, e.g. pattern-searching or attempting to improve), are attempting to please the experimenter, have a time-spending heuristic, or finally, suffer from a failure in the ability to reach acceptable accuracy within a reasonable time. Thus far, there is no evidence to support the Ability Failure hypothesis, the discovery of which was the main purpose of this design. However, this experiment also revealed several important results regarding the other potential causes.

First of all, I found that SDR was a significant mediator of the NFC-time relationship; this suggests that, at least in this sample, greater social desirability was in part responsible for the greater amount of time spent on the task. Of course, though, because both NFC and SDR are intra-psychic measures, it is impossible to determine the direction of the causal arrow between them (i.e., SDR may be the mediator, or NFC may be; the statistics do not tell us which is which). To speculate, it may be that

those with high SDR were more easily trained by others, due to their desire to please, and one common part of training students from a young age may be teaching them to spend time deliberating about problems (J. Baron, 1985). As students, they may have then discovered they enjoyed thinking hard and were rewarded for it, so they naturally developed high NFC. It's also possible that, again due to what they were taught, they simply think that others want them to have high NFC and therefore report it even when it is not entirely accurate. In any case, this discovery is an important clue to the reasons for the behaviour of high NFCs. The Social Desirability hypothesis certainly gains some support. It is also important to note that the mediation just discussed was not complete; i.e., while SDR explains some of the relationship between NFC and time taken, there is room for other explanations as well.

The improvement results, which showed that high NFCs were able to maintain their performance while low NFCs declined slightly, suggest that high NFCs are able to keep their focus on the task; this makes it likely that they are not simply bored and thinking of other things. Thus the results are not consistent with the Boredom hypothesis. Note that improvement results only pointed to a maintenance of performance and not actual improvement; they are not, then, strong support for the Improvement hypothesis. It is still possible, though, that high NFCs were learning to improve slightly, and this is what allowed them to maintain performance. Therefore, these results are consistent with Improvement but provide no real support for it.

The main effect of accuracy focus suggests that, even though it was not a significant mediator in this analysis, it is indeed significantly related to NFC and is worth investigating further. My measure of accuracy focus, as a self-report measure, may not be appropriate for detecting a true focus on accuracy. Therefore, the main focus of Experiment III is to examine participants' thresholds – that is, the strength of

evidence they require in order to stop sampling and make a choice. If high NFCs are truly more focused on accuracy, their sampling thresholds should be higher.

Finally, the result that high NFCs spend longer (in Block 1) on weak evidence trials, while low NFCs do not, speaks to the Accuracy Focus hypothesis. In particular, it suggests that high NFCs are sensitive to evidence strength, likely using it as an important factor in deciding when to stop sampling. Meanwhile, low NFCs could very well have a simple heuristic strategy, suggesting to them to stop sampling after a certain amount of time, regardless of how strongly the evidence points to one candidate or the other. Alternatively, their threshold is such that they will accept much weaker evidence in general, so that weak evidence and strong evidence trials both result in quick decisions. This is strong evidence in favour of the Accuracy Focus hypothesis.

To summarise, I have thus far found evidence in two experiments for a positive relationship between NFC and time spent gathering information when cognitive ability is controlled. Though the current experiment was designed to detect it, I have found no evidence that this relationship is due to an ability failure on the part of high NFCs. As for the other hypotheses, the data so far do not speak to the hypotheses of Improvement, Pattern-searching, or Time Heuristic. However, the result that high NFCs spend more time when evidence is weaker suggests they are more focused on accuracy, and the clear relationship between NFC and the accuracy focus self-report measure also supports this claim. Additionally, I have found a significant relationship between NFC and improvement (or, more aptly here, maintaining performance), suggesting that participants are not simply bored with the task. I also found that SDR was a significant mediator of the NFC-time relationship in Experiment II, suggesting that, at least in that sample, creating a positive impression

is part of the reason participants spend time. Thus overall, the reasons for the behaviour of high NFCs are beginning to become clear. The next experiments will focus on the potential desire for accuracy among high NFCs.

4. Experiment III: Threshold determination

Experiments I and II examined the participants' focus on accuracy, one potential reason for high NFCs' tendency to spend more time sampling evidence. Although I found no mediating role of my self-report measure in the NFC-time relationship, for either experiment, in each case it was significantly related to NFC, such that high NFC is related to a greater self-reported focus on accuracy. Further, I found that high NFCs in Experiment II spent more time when evidence was drawn from the weaker evidence universe, suggesting that they are aiming for a certain accuracy level, one they are willing to sample more to achieve. I designed Experiment III to be a stronger test of whether NFC is truly related to a focus on accuracy.

As I explained in Chapter 1, sampling theory typically describes the sampling-based choice problem as one of reaching a certain threshold of evidence strength (Busemeyer & Townsend, 1993; Merkle, Sieck, & Van Zandt, 2008; Pleskac & Busemeyer, 2009; Ratcliff & McKoon, 2008). Using this task as an example, imagine the sampler's view of Candidate A. At first, the proportion of smileys to frownies may be unclear, but by the law of large numbers, the more she samples, the closer the sample proportion gets to the population proportion. The same is true for Candidate B and, indeed, the difference between Candidates A and B, called the *contingency*. (To illustrate the contingency: Candidate A may have 80% smileys, while Candidate B has 60%, for a difference, or contingency, of 20% or .2.) At some point, a rational sampler will be happy enough that this contingency is not zero and in fact points to one candidate being superior – this is the sampler's threshold. Note that, although Fiedler and Kareev (2006) have argued that samplers are always more likely to accept higher contingencies, I have rebutted their argument (L. Evans & Buehner, in press), showing that both the magnitude of the contingency *and* the sample size matter. That

is, a contingency of 1 may not be convincing in samples of size 1 (i.e., 1 smiley for A, 1 frowny for B), but it may be very convincing in samples of size 10 (i.e., 10 smileys for A, 10 frownies for B).

It should be clear that, the more one is focused on accuracy, the longer one will sample in order to achieve more certainty about the proportions of smiley faces for each candidate (and therefore the contingency). However, as I discussed at length in the previous chapter, there may be several other reasons for sampling for longer: The sampler may be searching for patterns, distracted due to boredom, trying to please the researcher, or even at an ability disadvantage. A good measure of true accuracy focus can still be found, though: Look at the threshold of choice. I designed Experiment III such that participants would receive *fixed-size* samples; on each trial, they were then asked whether they would prefer, if they could, to continue sampling, or if they were happy to stop with what they had. When presented with all the evidence up front in this way, they are much less likely to be distracted, using a time-spending heuristic, or spending time to please the researcher, as they need spend much less time at the problem. Nor is there likely to be any ability differences on this problem, as there is no longer a need to remember the faces; they all remain on-screen. Participants may still do some pattern-seeking, but a desire to seek further (by indicating they would like more evidence) is likely to be attenuated by the fact that they are at a convenient stopping place. Thus determining on which trials they would prefer more evidence should be a good measure of their threshold. Imagine, for example, a trial of 8 faces per candidate, showing a contingency of .25. A person who is happy to accept this evidence can be said to be less focused on accuracy than someone who wishes to sample further, as the latter participant clearly wants more certainty that this difference is a true one.

I therefore varied not only the sample sizes participants saw (4, 8, 16, and 32 for each candidate), but the contingencies presented, such that each sample size was represented with a full range of contingencies, varying from 0 to 1. By determining which contingencies, on average, a participant is likely to accept for each sample size, I could determine whether she is more or less focused on accuracy. I predicted that high NFCs, in general, would show a greater accuracy focus.

4.1. Method

4.1.1. Participants

Participants were the same as those who participated in Experiment II. They were 30 adults (25 female and 5 male) recruited from the Cardiff University Human Participants Panel, who received course credit for their time. One participant did not complete the questionnaires and was removed from analysis. Of those who were later included in the analysis, the median age was 19.

4.1.2. Design

I presented participants with samples of 4, 8, 16, and 32 items of information for each of two candidates, *A* and *B*; I also varied the contingencies implemented by these samples (see Table 4.1). I asked participants to choose one or the other candidate as superior or, if uncertain (and desirous of more evidence), to make no choice. For each sample size there were 10 contingencies. For sample size 4, these were all possible non-zero contingency table configurations (ignoring the sign, which is irrelevant for my purposes here)[9]. For other sample sizes, I divided the contingency spectrum from 0 to 1 into eight ranges (0 to .125, .125 to .25, etc.), and participants

[9] A contingency of, say, +.50 is identical to one of –.50, except that rather than favoring option *A*, it would favor option *B*.

were presented with at least one contingency from each range. Because each sample size employed ten contingencies, the final two contingencies were determined randomly from all possible; in the case of sample size 8, this meant that two contingencies were repeated. When there was more than one possible evidence configuration that would create a particular contingency (e.g., 6/8 smileys for A and 4/8 smileys for B is one way to express a .25 contingency; another way would be 5/8 smileys for A and 3/8 smileys for B), I randomly determined which option was realized. I also always randomly determined whether a given contingency favored *A* or *B*. These determinations were done prior to the experiment, so that each participant would see each of these pre-chosen contingencies, although in a random order. Table 4.1 lists the contingency x sample size realizations presented to participants. Inspection of Table 4.1 shows that the average contingency shown was near .50 for each sample size, as was the outcome density *P*(O), the proportion of total positive information (smiley faces).

Table 4.1. Experiment III: Proportions of positive information and contingencies for each sample size

	Sample size											
	4			8			16			32		
	ϕ_A	ϕ_B	$\lvert\phi_A - \phi_B\rvert$	ϕ_A	ϕ_B	$\lvert\phi_A - \phi_B\rvert$	ϕ_A	ϕ_B	$\lvert\phi_A - \phi_B\rvert$	ϕ_A	ϕ_B	$\lvert\phi_A - \phi_B\rvert$
	.00	.25	.25	.00	.13	.13	.44	.50	.06	.97	1.00	.03
	.25	.50	.25	.88	.75	.13	.75	.94	.19	.34	.41	.06
	.50	.75	.25	.38	.13	.25	.06	.38	.31	.41	.19	.22
	.75	1.00	.25	.50	.13	.38	.50	.19	.31	.44	.72	.28
	.00	.50	.50	.38	.88	.50	.38	.81	.44	.44	.00	.44
	.75	.25	.50	.75	.13	.63	.94	.38	.56	.66	.06	.59
	1.00	.50	.50	1.00	.38	.63	.31	.94	.63	.22	.84	.63
	.00	.75	.75	1.00	.25	.75	.88	.06	.81	.94	.13	.81
	1.00	.25	.75	1.00	.13	.88	.94	.06	.88	.03	.94	.91
	1.00	.00	1.00	.00	1.00	1.00	1.00	.00	1.00	1.00	.00	1.00
Average			.50			.53			.52			.50
P(O)			.50			.49			.52			.49

Note. For each sample size, displays all ϕ_A, the proportion of positive information for *A*, ϕ_B, the proportion of positive information for *B*, and the observed contingency, $\Delta_O = |\phi_A - \phi_B|$, that the participants saw. I also report, for each size, the average Δ_O and the average *P*(O), the proportion of total positive information seen for both candidates on a given trial.

4.1.3. Procedure

As in Experiments I and II, participants were always tested alone, with the experimenter sitting in the next room. They entered and were seated at a computer, which provided both the bulk of the instructions and the tasks. In this case, participants first performed the self-truncated sampling task that made up the task portion of Experiment II.

After having completed this task, participants then proceeded to the threshold testing task, which was introduced as a different approach to the same problem of determining the better of two job candidates. Participants were informed that from now on, and in contrast to the previous task, they would be presented with a fixed amount of information on each candidate, and that they had to indicate which of the two was superior; if they felt they could not make a sound decision based on the information presented, however, they could choose to "set aside" the current data to return to it and gather more later (although they would not actually do so). Unlike in Experiments I and II, participants were encouraged to be *accurate* in their judgments; this was to ensure that they at least sometimes chose to gather more data, as Fiedler and Kareev (2006) reported that in a similar task participants almost always made a choice between candidates, even with very small sample sizes.

Contingency information was again presented in the form of smiley and frowny faces, representing positive and negative information, respectively, for each candidate *A* and *B*. Information for *A* was presented on the left half of the screen, which was colored turquoise, and information for *B* was presented on the right half of the screen, which was colored orange. Smiley faces were presented in red and frowny faces in black, and like faces were always grouped together to minimize the scope for misinterpretation of the evidence (see Figure 4.1). Below the contingency

information, participants saw three options: Choose candidate A, Choose candidate B, or No Choice (Gather More Data). They were then asked to report their confidence on a scale of 1 to 21, using a sliding scale bar as in Experiments I and II. Once participants decided on an option for a problem and reported confidence, a white screen was presented for two seconds, and they were then presented with the next choice problem.

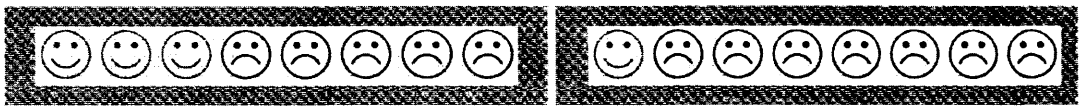


Figure 4.1, Experiment III. Screenshot of one of the contingencies used.

4.1.4. Materials

Because this experiment was run concurrently with Experiment II, the Cognitive Ability, NFC, and SDR materials, their mean scores and Cronbach's α, were naturally identical to those in Experiment II. To repeat them: Mean scores were –1.14 (*SD* = 6.88) for the Raven matrices, 86.10 (*SD* = 12.12) for the NFC measure and 16.07 (*SD* = 5.73) for the SDR measure; these means did not differ significantly from those in Experiment I. Cronbach's α for these measures were .71, .89, and .74 respectively.

4.2. Results

I removed the same participant from analysis as in Experiment II, since this participant sample was the same. This again resulted in 28 participants included in the final analysis.

Because this experiment did not involve sampling from populations in order to present the judgment evidence, there was no "correct" answer for participants to

choose on a given trial. However, all participants, when making a choice, always chose the candidate with a greater proportion of smiley faces, with only one participant making a single trial exception.

Participants' Median Confidence (out of 21) in this experiment was again calculated only for those trials on which the participant made a choice. I also measured the *average* contingency for which participants were happy to make a choice, for each sample size. I chose to average these rather than simply take the minimum contingency each was willing to accept (which might seem like the appropriate measure of a minimum threshold) because participants were not always consistent in their choices. E.g., a participant might sometimes make a choice with a contingency of .5, for size 4, but sometimes refrain on a .5 trial and prefer to gather more evidence. Thus averaging, while not a true measure of their minimum threshold, is an overall better measure of their general preference.

Because I intended again to include Cognitive Ability as a covariate in the main analysis, I relied on my analysis from Experiment II, which indicated it was an appropriate covariate to use, as it does not appear to be significantly related to NFC in this sample and therefore should exhibit a unique effect on any outcome variable. Cognitive Ability was not correlated significantly with Median Confidence at any size (all rs < .28 in magnitude), nor with participants' Mean Contingency score at sizes 8, 16, or 32 (all rs < .20 in magnitude), although it was correlated with Mean Contingency at size 4, $r = .42$, $p = .03$, such that a higher Raven score indicated a greater likelihood of requiring a higher contingency to make a choice. Finally, Cognitive Ability again did not contribute significantly to any of the following analyses (all Fs < 2.33).

4.2.1. Main analyses

I first examined the effect of NFC on participants' Mean Contingency scores using a 4 (Sample Size: 4, 8, 16, or 32) x NFC (measured)[10] ANCOVA with Mean Contingency as the dependent measure, including Cognitive Ability as a covariate. Figure 4.2 displays the results. It is first of all clear that, as expected, participants' Mean Contingency required for choice declined with Sample Size; this, again, is due to the law of large numbers, such that smaller contingencies are often acceptable with larger sample sizes because one is more confident that they represent the population contingency (L. Evans & Buehner, in press). This main effect of size was significant,

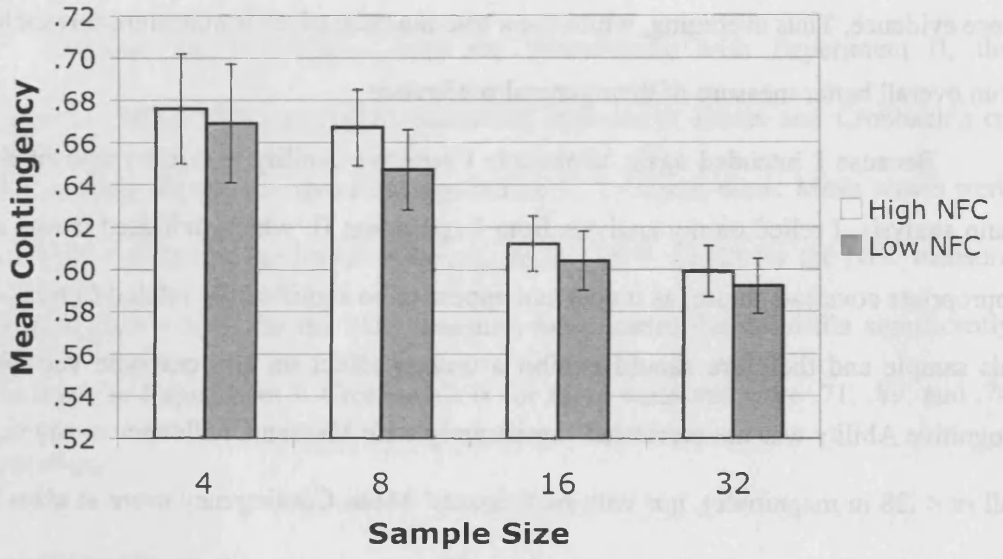


Figure 4.2, Experiment III. Average Mean Contingency for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), for sample sizes of 4, 8, 16, and 32 pieces of evidence for each candidate. Mean Contingency indicates the average contingency a participant was willing to use in order to make a choice on a given trial. Bars are standard 1 error.

[10] Again, all analyses were also run using the median split of NFC (*Median* = 84.0), with a similar pattern of results. In this case, though, the NFC x Sample Size interaction in the Median Confidence ANCOVA was not significant, $p = .16$.

$F(1.97, 49.35) = 10.46$, $p < .001$, $\eta_p^2 = .295$. Second, although it is apparent from the graph that those high in NFC (estimated at $M + .5SD$) consistently required higher Mean Contingencies to make choices than those low in NFC (estimated at $M - .5SD$), this effect was not significant, $F(1, 25) = 0.56$, $p = .46$. There were no other significant patterns (all Fs < 1.14). Thus I did not find the predicted effect of NFC; high NFCs in this sample were not significantly more likely to require stronger evidence than low NFCs.

Second, I examined the effect of NFC on Median Confidence, using an identical ANCOVA analysis to the above, with Median Confidence as the dependent measure. It is clear from Figure 4.3 that, unsurprisingly, participants' confidence increased significantly with sample size, $F(3, 75) = 30.31$, $p < .001$, $\eta_p^2 = .548$. Clearly, as participants are happy with lower contingencies for larger sample sizes,

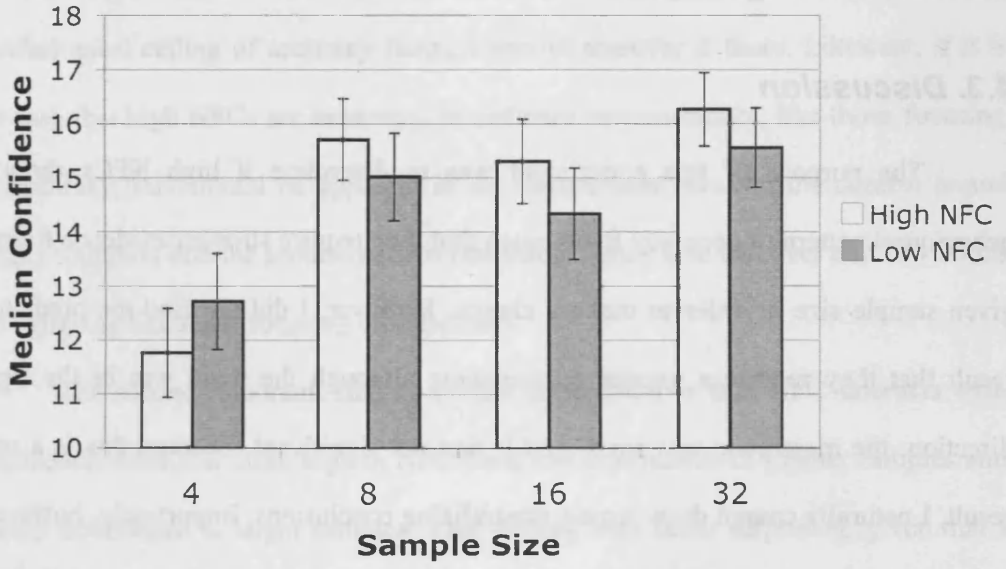


Figure 4.3, Experiment III. Average Median Confidence, on a scale of 1 to 21, for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), for sample sizes of 4, 8, 16, and 32 pieces of evidence for each candidate. Bars are 1 standard error.

and the full range of contingencies was presented, there were simply a greater number of large sample-size trials on which they could make a confident choice (some trials showing enormous differences in candidates). It is also apparent from the graph that high NFC participants' confidence was lesser than low NFCs' for the smallest sample size, size 4, but greater for each of the larger sample sizes; this interaction was also significant, $F(3, 75) = 17.59, p < .001, \eta_p^2 = .222$. Simple effects analyses, done using the median split of NFC (*Median* = 84.0), revealed that Sample Size was a highly significant predictor of confidence for both low NFCs, $F(3, 23) = 7.66, p = .001, \eta_p^2 = .500$, and high NFCs, $F(3, 23) = 16.74, p < .001, \eta_p^2 = .686$, the effect sizes showing that more variance is explained by Sample Size for those high in NFC. There were no other significant effects (all Fs < 2.33). Thus it is clear both that Sample Size in general has a significant effect on people's confidence in their choices, and it appears that this effect is differentially based on NFC, such that those high in NFC are overall more confident in large samples (8, 16, 32) and less confident in small samples (4).

4.3. Discussion

The purpose of this experiment was to determine if high NFCs show a behavioural pattern of accuracy focus, such that they require stronger evidence from a given sample size in order to make a choice. However, I did not find the predicted result that they require a greater contingency; although the trend was in the right direction, the magnitude was small and it was not significant. Because this is a null result, I naturally cannot draw strong generalizing conclusions. Importantly, however, the participants used in this experiment were the same as those in Experiment II – for which I found a significant effect of NFC on time taken, with cognitive ability

removed. It appears that these same participants have not indicated any preference for stronger evidence within a given sample size, compared with their low-NFC peers.

One possible criticism of this study is that I asked participants to be as accurate as possible in their judgments, in order to elicit at least some cases in which they were willing to indicate they would not make a choice and gather more data. It could be argued, then, that everyone was so focused on accuracy that there should be no difference among the high and low NFCs. To counter this, I note that there was always an entire contingency range to choose from (0 to 1), and participants, on average, chose to make choices in the .6 to .7 range – thus there was plenty of room for more caution on the part of all participants. Still, perhaps they hit an equal psychological ceiling for evidence strength, and this may be the reason for the lack of difference. I will examine accuracy focus once more, then, in Experiment V, in which I asked some participants to focus on accuracy, others to focus on efficiency, and a third group to focus on them equally. If it is the case that all participants move to a psychological ceiling of accuracy focus, I should discover it there. Likewise, if it is the case that high NFCs are behaving, in ordinary circumstances, like those focusing on accuracy, this should be apparent in the comparison between the control (equal focus) condition and the accuracy focus condition. I may also discover that low NFCs are behaving like those focusing on efficiency.

The second important finding in this experiment is that NFC interacts with confidence, such that those high in NFC have less confidence in smaller samples and greater confidence in larger samples. This finding may seem surprising, given that I found no relationship between NFC and confidence in either Experiment I or Experiment II – and, as I mentioned, the participants from Experiment II were the very same ones who completed this task. However, my measurement of confidence in

Experiments I and II was of *median* confidence; i.e., how confident participants were on average throughout the trials. A measure of confidence split by sample size would have been difficult to ascertain in Experiments I and II, given that they involved self-truncation and therefore participants low in confidence could simply continue sampling until they were more confident. However, the current experiment provided a natural opportunity for such measurement. Despite the option to "gather more data," clearly the range of data participants were willing to accept allowed for their differential preferences to become apparent.

These preferences appear entirely consistent with the participants' behaviour in earlier experiments; high NFCs preferred to take longer (and therefore collect larger samples) in Experiments I and II, and here they indicate again a sort of preference for larger samples. However, it is of note that they exhibit a larger range of confidence in their choices than low NFCs, given that their contingency choices were similar. Perhaps they are simply more finely attuned to the task, more carefully selecting their confidence, adjusting it further from a psychological midpoint in both high and low cases. Perhaps, then, they even spend more time on judging their confidence. This would suggest that their time-spending habits generalize to internal judgment tasks. If they do in fact spend this extra time judging confidence, this would provide support for either the Accuracy Focus, Social Desirability, or Time Heuristic hypotheses, as it is unlikely that participants would be doing so in order to search for patterns, to improve, or out of boredom. A better task that is small in timescale will come in Experiment VI, in which the entire focus of the task is to answer a question within a small time window; I will time participants here and look to see if the NFC-time relation exists at this small scale.

In the next chapter, with Experiment IV, I examine the confidence growth of both high and low NFCs as they complete the task; this experiment should also shed more light on the confidence result of the present experiment, showing whether high NFCs are in fact typically less confident with smaller samples as their confidence grows.

The main conclusion from this study, then, is that, in this case, the thresholds of those high and low in NFC are not significantly different, when both are instructed to be as accurate as possible. Although, I found no significant evidence here to support or refute my hypotheses, I have thus far discovered evidence supporting the Social Desirability and Accuracy Focus hypotheses, and I have cast serious doubt on Ability Failure, and some doubt on the Boredom hypothesis.

5. Experiment IV: The accumulation of confidence

Experiments I, II, and III all examined the confidence of participants after they made their choices. For the past 100 years, this post-task confidence judgment has been the rule in psychological testing of confidence (Baranski & Petrusic, 1998), with very few exceptions. The purpose of Experiment IV was to determine confidence *during* the information-gathering stage of choice making.

The typical result of the past experiments, using end-of-task judgments of confidence, is exemplified in a study by Vickers, Burt, Smith, and Brown (1985). They collected confidence judgments under two conditions: one in which the experimenter controls the sample size, and one in which the participant controls it. They found that, in the former case, confidence reports rose with sample size, and in the latter case, confidence decreased with sample size. The reason for this discrepancy is clear: When the participant controls the sample size, she stops as soon as she encounters clear evidence. Thus small samples in the participant-controlled condition indicate that she has encountered particularly clear evidence and therefore stopped early; large samples indicate that the sample was mixed, and likely less clear – perhaps she even chose to give up and guess. Meanwhile, in the experimenter-controlled condition, the law of large numbers is in effect. Because the experimenter does not pay attention to anything but the sample size, the samples' evidence value is randomly determined, and in this case a large sample is always more likely to display stronger evidence. Similar results were reported when confidence was tracked during the experiment (Irwin, Smith, & Mayfield, 1956): Confidence rose with (experimenter-controlled) sample size.

However, no one has yet examined confidence growth during a participant-controlled scenario. Presumably this is because it is a difficult task to accomplish, as

the participant must not only determine the sample size but simultaneously provide the confidence reports throughout. The primary value of such a study is not to determine whether confidence will rise with sample size – there are strong theoretical reasons to believe that it will – but rather to see the shape of its growth curve. It may adhere to the typical and long-known learning curve (Bills, 1934; Ebbinghaus, 1885), or it may be altogether different, reflecting the nature of a measure that will fluctuate with evidence seen and may, for example, gain more from positive evidence than it loses from negative.

There are, of course, various theories of confidence that might be informed by this experiment (for a review, see Baranski & Petrusic, 1998). However, that discussion is beyond the scope of this thesis. My concern here is whether there may be individual differences in this confidence curve, such that individuals high in NFC differ from those low in NFC. Given the results of Experiment III, which showed that high NFCs report higher confidence in large samples and lower confidence in small samples than low NFCs, there is a good possibility that they will differ in their curves, when confidence is measured in real time. In this case, I would predict that high NFCs grow in confidence much more slowly at first, when sample sizes are small, and may then change confidence sharply as sample size grows. Importantly, I might also find clues regarding the tendency of high NFCs to take longer to sample than low NFCs, when cognitive ability is controlled. In particular, a shallower confidence curve overall could indicate a cautious nature (and therefore accuracy focus) or distraction (by daydreaming or task-related thinking, which causes difficulty in remembering the presented information and therefore the lower confidence). Meanwhile, a similar confidence curve may indicate that high NFCs simply linger at high confidence, unable to disengage from the task – this may be due either to simple difficulty with

disengagement (as Time Heuristic predicts) or a desire to spend time to please the researcher.

5.1. Method

5.1.1. Participants

Participants were 37 adults (32 female and 5 male) recruited from the Cardiff University Human Participants Panel, who received course credit for their time. Their median age was 19.

5.1.2. Procedure

Participants were tested in groups of two to four in a small laboratory with screen dividers in between computers. Before the experiment began, the participants were gathered in front of one computer for a demonstration of the task, where the experimenter gave a verbal summary of the written instructions they would see upon starting (see Appendix D), in addition to demonstrating the features of the interface that would allow the participants to begin and end the trial (explained below). Once the choice task was finished, the participants completed the Raven matrices and the questionnaires.

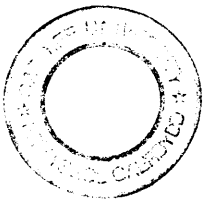
5.1.3. Choice task

The choice task presented to participants was identical to that in Experiments I and II, with two exceptions. First, although in previous experiments the purpose of mixing strong and weak evidence universes was purely for variety, in this experiment I planned to examine confidence within different universes; this was in case confidence might be, for example, at ceiling in a strong evidence universe and/or at floor in a weak evidence universe. Thus, to increase the chances of employing a

universe that would result meaningful amount of variance in confidence, in this experiment the number of universes of evidence was increased to three. I included both the strong and weak evidence universes from Experiments I and II, in addition to a "moderate" universe, which was comprised of 60% positive judgments and 40% negative judgments for the better candidate, and 40% positive and 60% negative for the worse. All participants saw 8 trials from each universe, randomly intermixed.

Second, participants had a method for tracking their confidence on-line. The screen included a small slider bar, which began with its button on the far left. Any movement of this slider bar would begin the trial, and evidence would begin to appear for the candidates. Participants were instructed that movement of this slider to the right would indicate increased confidence, and reaching the far right-hand end would indicate enough confidence to make a choice and would stop the trial, which would present them with the candidate options ("Choose Candidate A," "Choose Candidate B," or "No Choice"). They were explicitly informed that the only other purpose of the bar was to keep track of their confidence during the trial, whether it increased or decreased. The trial could also be stopped at any time by pressing a button located just below the slider bar, which read "Stop Now."

Confidence was recorded based on the position of the slider bar, on a scale of 0 to 100; the far left was scored as 0 and the far right was scored as 100. Every time a smiley or frowny face appeared on screen, the position of the confidence slider was recorded. In order to allow more time for participants to respond with the slider bar, faces remained on screen for 1200 ms, instead of the 500 ms allowed in Experiments I and II.



5.1.4. Materials

The Cognitive Ability, NFC, and SDR materials in this experiment were identical to those in Experiments I, II, and III. Mean scores were –1.29 (*SD* = 7.76) for the Raven matrices, 74.68 (*SD* = 14.28) for the NFC measure and 18.50 (*SD* = 4.40) for the SDR measure. Cronbach's α for these measures were .78, .87, and .52 respectively.

Of note, the NFC and SDR measures both had significantly different means across the three samples in Experiments I, II, and IV, $H(2) = 13.69$, $p = .001$, and $H(2) = 6.28$, $p = .04$, respectively; there was no difference in Cognitive Ability, $H(2) = 0.26$. Post-hoc tests, with a Bonferroni-corrected alpha of .017, revealed that, in the case of NFC, the current experimental sample had a significantly lower mean than in Experiments I and II, $U = 272.5$, $z = -2.88$, p = .004, and $U = 238.0$, $z = -3.37$, p = .001, respectively. For SDR, post-hoc tests were not significant at the reduced alpha, although the nearest to significance was between Experiments II and IV, $U = 318.5$, $z = -2.23$, p = .03, such that the current experiment's participants scored higher than those in Experiment II. The significance of these findings will be discussed later.

5.2. Results

There were two participants who spent extraordinary time at this task – with values 3.7 and 2.8 standard deviations from the mean. Because together their scores exerted considerable influence on the mean time spent, I decided to consider whether they might be outliers when their influence was removed from the sample. I thus removed them and calculated the mean and standard deviation of the distribution without their inclusion. When I calculated this adjusted mean, they were each greater than 4 standard deviations from it, and I therefore removed them from analysis. One

participant was also removed from analysis for scoring –10 in Accuracy (6 correct, 16 incorrect, and 2 no choices), indicating an error in carrying out the instructions. Thus, the final number of participants entered into analysis was 34.

Most dependent measures (Time Taken, Accuracy, and Improvement)[11] were recorded in the same way as in Experiments I-III, with the exception of Confidence, which was recorded by the position of the slider bar throughout the task (detailed above). The average amount of (log) time spent during this experiment was 3.13 (*SD* = 0.12). I chose not to compare Time Taken directly to other experiments, as faces were presented for a different length of time in this experiment (1200 ms each instead of 500 ms); in its stead, I compared participants' Median Sample Size, or the average amount they sampled. Performing a Kruskal-Wallis test, I found that Median Sample Size was significantly different across Experiments I, II, and IV, $H(2) = 26.99$, $p < .001$. Post-hoc tests, conducted at a Bonferroni-corrected alpha of .017, revealed that participants in the current experiment sampled significantly more ($M = 29.14$, $SD = 11.10$) than in either Experiment I ($M = 21.26$, $SD = 7.96$), $U = 283.50$, $z = -2.72$, $p = .006$, or II ($M = 15.25$, $SD = 6.90$), $U = 127.0$, $z = -4.94$, $p < .001$.

In order to compare Accuracy to previous experiments, in which there were fewer trials (20 instead of 24), I adjusted the scores in the current experiment. I removed an amount equal to the average percentage of correct trials times 4, and added an amount equal to the average percentage of incorrect trials times 4; this would, on average, remove the extra points gained in the final 4 trials and add the extra points lost. This resulted in an overall mean Accuracy of 12.09 (*SD* = .4.02). I then performed a Kruskal-Wallis test, with Experiment as a factor, finding that Accuracy was significantly different across experiments, $H(2) = 12.26$, $p = .002$. Post-

[11] AE Focus was excluded from this task, as it was not the principle focus of this experiment.

hoc tests, at a Bonferroni-corrected alpha of .017, revealed that in the current experiment, participants were significantly more accurate (M = 12.09, SD = 4.02) than in Experiments I (M = 8.54, SD = 3.44), U = 242.0, z = –3.32, p = .001, and II (M = 9.57, SD = 3.35), U = 304.0, z = –2.44, p = .015. The mean score for Improvement was –.004 (SD = .23), and there was no significant difference in Improvement across experiments, $H(2) = .26$, $p = .88$.

Taken together with the differences I reported earlier in participants' NFC and SDR scores, the most likely possibility is that all of these discovered differences are a result of the major change in this experiment in comparison to Experiments I and II: the focus on tracking confidence during the task. Given the extra cognitive load of keeping track of one's confidence during each trial, it is not surprising that participants would sample more – likely due to proceeding more slowly in general – and, as a consequence, achieve higher accuracy. It is, however, surprising that they would subsequently report lower NFC and (marginally) higher SDR. Perhaps the sample of participants in this experiment was simply different from the others, due to random fluctuation. There is also a slight chance that these differences arose from their extra effort; although NFC is typically conceptualised as a stable trait (Cacioppo & Petty, 1982), Baron (1985) has hypothesized that it may be malleable. However, it is surprising that they should expend so much effort and then report less enjoyment of it. There is little that can be definitively concluded here without research that specifically tests the hypothesis of malleability.

Also in line with my interpretation of extra cognitive effort required, I found that Cognitive Ability was significantly correlated with Accuracy, $r = .46$, $p = .006$. Participants with higher Raven scores typically have a higher working memory capacity (Conway et al., 2002; Engle et al., 1999; Kane et al., 2004), and it is

therefore very likely they would be better able to focus on both tracking confidence and remembering the faces.

Given the difficulty of altering the task to be less cognitively draining, and given that there was still a potential role for NFC in both time taken and confidence curve, I chose to proceed with the analysis.

5.2.1. Main analyses

I first examined the effect of NFC on Time Taken with a 3 (Universe: Weak, Moderate, or Strong) x NFC (measured) ANCOVA, with Cognitive Ability entered as a covariate. In this case there was no significant effect of either NFC or Cognitive Ability (both Fs < 1.15), although there was the typical main effect of universe, F = 3.32, p = .04, η_p^2 = .097, such that the weak evidence universe took the most time (M = 2.69, SE = 0.02), followed by the moderate universe (M = 2.66, SE = 0.02), followed by the strong universe (M = 2.59, SE = 0.02).

There was similarly no effect of either NFC or Cognitive Ability in a simultaneous regression of the two variables onto Improvement (both ts < 0.14 in magnitude). Also, in a second, identical regression, there was no effect of NFC on Accuracy, although there was an effect of Cognitive Ability, such that it explained 23.6% of the variance uniquely (see Table 5.1). As discussed above, this effect is likely the result of the increased cognitive load during this task, such that those high in Cognitive Ability were able to better remember faces while simultaneously tracking their confidence.

Table 5.1

Experiment IV: Simultaneous regression of NFC and Cognitive Ability onto Accuracy

	β Weight	t Value		Unique variance explained	Partial r
Criterion variable: Accuracy					
Cognitive Ability	.503	3.097	**	.236	−.468
Need for Cognition	−.149	−.920		.021	.403
Overall regression:					
F = 4.81*					
Multiple R = .486					
Multiple R-squared = .237					

Note. β weights are standardised.

*$p < .05$, **$p < .01$.

5.2.1.1. Confidence

To begin my analysis of confidence, I chose to examine the growth of confidence during the first 15 faces seen. The reason I did not examine confidence across *all* faces seen is because evidence sampling varied widely, both across trials and across participants. Thus to examine confidence up to the maximum sample size reached would have entailed comparing, for example, 50- and 70-sample trials with 10- and 15-sample trials, which would naturally lead to difficulties, as the latter trials reached maximum confidence before the former were even half finished. The scope of their curves would naturally be greatly different. Considering that 81% of trials included 15 faces or more, this amount provided a reasonable level of comparison. For those 19% of trials on which 15 faces were not reached, I substituted confidence from the end of the trial up to 15 as follows: If the trial finished with a choice made, I substituted 100 (maximum) confidence thereafter, and if no choice was made, I substituted 0 (no confidence). E.g., if the trial ended after 12 faces with a choice, I

assigned 100 confidence for faces 13-15. As in previous experiments, the number of no-choice decisions was low; only 8.3% of trials resulted in no choice.

Finally, in order to make the subsequent analysis less unwieldy, I chose to look at the median confidence for each group of three faces: 1-3, 4-6, 7-9, 10-12, and 13-15. Thus, for a given participant, I took the median confidence of, e.g., faces 1-3 over all trials. Median Confidence was not normally distributed for each of these levels, so I performed square root transformations. I also removed the 1-3 and 4-6 levels from analysis, which were the most problematic and also the least likely to show NFC-based differences, since participants would all be just beginning to adjust to the trial. See Figure 5.1 for a graph of the median values for each of the five sample-size levels.

Because I intended to conduct an ANCOVA that included Cognitive Ability as a covariate, I examined the correlation between Cognitive Ability and NFC (the only

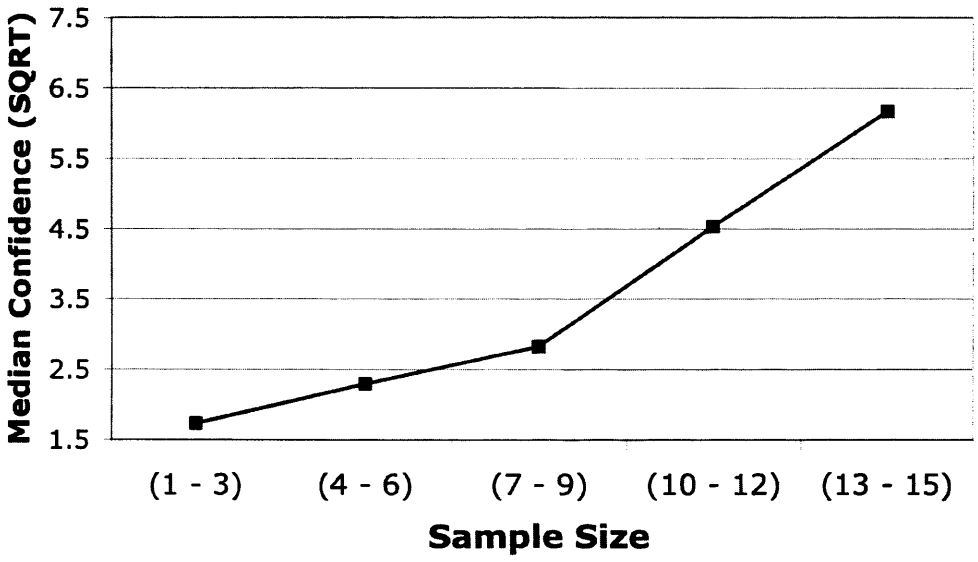


Figure 5.1, Experiment IV. Median of Median Confidence, collapsed across NFC and evidence universes, at sample sizes of 1-3, 4-6, 7-9, 10-12, and 13-15 pieces of evidence for each candidate.

between-subjects measure) in order to determine if they were independent. This correlation was nonsignificant, $r = .14$, $p = .45$, indicating that Cognitive Ability was appropriate to include. Cognitive Ability was not correlated with Median Confidence at any level of sample size or evidence strength (all rs < .21 in magnitude), and it did not play a significant role in the following analysis (all Fs < 0.33).

To determine the shape of confidence, for high and low NFCs and as sample size increased, I examined the effect of NFC on participants' Median Confidence using a 3 (Sample Size: 7-9, 10-12, 13-15) x 3 (Universe: Weak, Moderate, or Strong) x NFC (measured)[12] ANCOVA, with Median Confidence as the dependent measure and including Cognitive Ability as a covariate. See Figure 5.2 for the results.

It is apparent and not surprising that participants were more confident as sample size increased, $F(1.37, 42.56) = 72.06$, $p < .001$, $\eta_p^2 = .699$, and that they showed greater confidence the stronger the evidence universe, $F(2, 124) = 5.98$, $p = .004$, $\eta_p^2 = .162$. The interaction between sample size and evidence strength was marginal, $F(2.92, 90.54) = 2.71$, $p = .051$, $\eta_p^2 = .080$. It is clear from the graphs that, as sample size increased, evidence strength was more likely to play a role in confidence; simple effects analyses showed that evidence strength, while not exhibiting a significant effect in the sample size range 7-9, did play a significant role in the ranges of 10-12, $F(2, 30) = 6.07$, $p = .006$, $\eta_p^2 = .288$, and 13-15, $F(2, 30) = 3.35$, $p = .049$, $\eta_p^2 = .183$. I.e., the evidence universes became distinguishable from each other at 10-12 pieces of evidence.

[12] I performed an identical analysis with a median split of NFC (*Median* = 75.0), with a similar pattern of results.

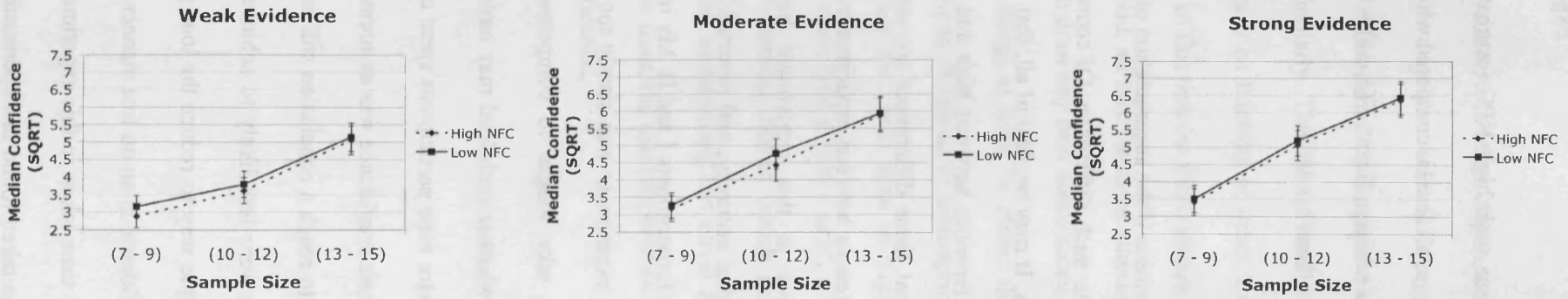


Figure 5.2, Experiment IV. Average Median Confidence (square-root transformed) for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), at sample sizes of 7-9, 10-12, and 13-15 pieces of evidence for each candidate. The weak evidence universe consisted of a population contingency .1, moderate evidence .2, and strong evidence .4. Bars are 1 standard error.

Although it is apparent in the graphs that those with high NFC (estimated at $M + .5SD$) did report consistently lower confidence, at all levels, compared with those low in NFC (estimated at $M - .5SD$), this trend was not significant, $F(1, 31) = 0.26$. There were no other significant trends in the data; all other Fs < 0.33.

5.3. Discussion

The aim of this experiment was to determine if the accumulation of confidence differs among high and low NFCs. I found no such difference. Of course, this null result could support any number of hypotheses. It may be, first of all, that the result is a Type II error, such that there are truly differences between high and low NFCs, although the small standard errors suggest that these differences are unlikely to be great. It may also be that the discovered differences between experiments is the cause of the result. I discovered that participants in this experiment both behaved differently, sampling more and achieving higher accuracy, and reported NFC scores that were significantly different from those in Experiments I and II. My interpretation of the differences I found is that the current experiment was simply too cognitively demanding, such that participants had to take longer to compensate (gaining somewhat in accuracy). It is difficult to tell whether such load may have influenced reports of confidence. It's possible that the extra time participants spent compensated for the difficulty of the load, and therefore their confidence was as normal. It is also possible that this feeling of expending effort to reach a conclusion still had an effect, (falsely) suggesting to participants that they were less likely to achieve the correct answer. It is unfortunate that there is no simple way to reduce the load generated in this experiment and yet continue to track confidence in an on-line manner.

Second, I also examined participants' time taken in this experiment to search for a replication of the effect of NFC on time taken found in Experiments I and II. I

did not find such a replication. Again, it is difficult to determine why this might be the case, although cognitive load is again a possible candidate. It may well be that participants, all burdened by the need to devote extra attention to confidence tracking, performed similarly. This interpretation is certainly in line with my hypotheses that high NFCs may be thinking of other things (task related or not) during the ordinary instantiation of this task (in Experiments I and II) – now that everyone is distracted by extraneous load, all take the same amount of time. However, given the null result, it is impossible to draw any firm conclusions.

One thing, at least, is clear: the overall shape of confidence as it grows. Although, due to violations of assumptions at very small sample sizes, I was not able to test whether the curve seen in Figure 5.1 is significantly nonlinear, its shape suggests the same type of curve as x^n, a slow, flat portion that then steeply rises. This is typical of learning curves, which generally follow an S shape (Fitts & Posner, 1967). It is an unsurprising but novel result, suggestive of the participant's role in learning, over time, the true proportions of evidence for each candidate, and therefore her true confidence.

6. Experiment V: A further test for accuracy focus

One of my primary hypotheses about the relationship between NFC and information gathering is that those high in NFC spend more time gathering evidence in order to pursue greater accuracy (even if they do not succeed, or succeed only marginally). In Experiments I and II, I discovered that NFC was significantly correlated with a self-reported measure of accuracy focus, although this measure did not significantly mediate the relationship between NFC and time spent in either experiment. In Experiment II, I found that only high NFCs tended to spend more time on weak-evidence trials, compared to strong-evidence trials, suggesting that they were sensitive to the strength of evidence, and thus accuracy was important to them. In Experiment III, I examined whether high NFCs might require stronger evidence, for a given sample size, than low NFCs, finding only a slight trend that was not significant. Finally, in Experiment IV, I would have expected to find a shallower confidence curve for high NFCs if they were more focused on accuracy, and I did not. Thus, the evidence so far is mixed; high NFCs report that they value accuracy, and clearly differentiate between weak and strong evidence, but this valuation of accuracy is not reflected in the mediation analysis, nor in their apparent (lack of) preference for stronger evidence, nor in their confidence curve.

However, there are three reasons to continue to search for evidence that accuracy focus may be the reason that high NFCs spend more time gathering evidence. The first is that this hypothesis is the most likely. It is both simple – participants clearly understand that sampling for longer will lead to greater accuracy – and in line with the direct reports of participants. The second is that the picture painted by Experiments I-IV is incomplete. In Experiment III, participants may have shown little difference between their evidence strength preferences because they were

all very focused on accuracy. In Experiment IV, the high cognitive load of the task may have affected confidence. And Experiments I and II, though they showed strong effects, were conducted on relatively small sample sizes (final $N = 28$ in each case), so I may have lost the power needed to detect mediation. And third, the evidence from Experiment II that it is only the high NFC participants who spend more time when evidence is weak is strong behavioural evidence that their focus on accuracy is greater.

I therefore chose to continue to examine this hypothesis, and, as an efficient way to address the issues I encountered in Experiments I-IV, to conduct a larger study, one in which participants were instructed to either focus on accuracy (Accuracy condition), focus on speed (Speed condition), or focus on both equally (Control condition). The latter condition served as a replication of Experiments I and II with a larger sample size, in order to hopefully discover any smaller effects, while the former two conditions could be compared to this control condition. If high NFCs are focusing on accuracy, I would expect that high NFCs in the Control condition would behave similarly to those instructed to focus on accuracy – i.e., they should take a similar amount of time to sample evidence. If low NFCs are focusing on speed, I should see similarities between low NFCs' performance in the Control group and the performance of the Speed group.

Finally, such an experiment gives me another opportunity to search for an ability failure of high NFCs, which I originally did in Experiment II. In that experiment, in order to test for ability failure in both a between- and within-subjects manner, participants began the experiment with a block in which everyone received the control instructions to focus equally on both accuracy and efficiency. However, this gave them some time to practice at the task, which may have ameliorated any

deficit in ability by the time they reached Block 2. In this experiment, all participants begin with separate instructions (in particular, Control or Speed). Thus if there are any ability differences, they are more likely to be apparent here. If high NFCs are suffering from a poorer ability to remember the faces, their performance in the Speed condition should be worse than that in either of the other two conditions.

6.1. Method

6.1.1. Participants

172 adults (160 female and 12 male) were recruited from the Cardiff University Human Participants Panel, with median age 19. They received course credit for their involvement. In total, there were 59 participants in the Control condition, 59 in the Accuracy condition, and 54 in the Speed condition.

6.1.2. Procedure

Participants were tested together in a large laboratory with seating for up to 16 people, with dividers set up in between each of the computers. All participants were tested in groups of 10-15 at a time, with the exception of one group of 4, and every time such a group was tested all members were part of the same condition. As in Experiments I-III, instructions were delivered by the participant's computer. Once the choice task was completed, participants completed a task for a separate experiment (Experiment VI), which was followed by the questionnaires. When these were finished, the instructions requested the participant read a free novel online (Jane Austen's *Pride and Prejudice*); this was in order to fill time until all participants had finished, so that any participants still completing the experiment would not be

disturbed by the leaving of their fellows. Participants returned between one and seven days later in order to complete the Raven matrices; this was done in groups of 1 to 3.

6.1.3. Choice task

The choice task was identical to that in Experiments I and II. The only exception was that participants in the Accuracy condition received instructions to focus on accuracy and participants in the Speed condition received instructions to focus on efficiency (see Appendix E). Again, the word "efficiency" was used rather than "speed" in order to imply a minimum level of accuracy was required.

6.1.4. Materials

The Raven matrices, NFC and SDR questionnaires used in this experiment were identical to those in Experiments I-IV. Mean scores were –1.51 (SD = 7.70) for the Raven matrices, 78.95 (SD = 14.25) for NFC, and 16.61 (SD = 4.82) for SDR. Cronbach's α for these measures were .76, .93, and .66, respectively.

I compared the questionnaire measures in this experiment to those in Experiments I, II/III, and IV. SDR was significantly different across experiments, $H(2) = 8.46$, $p = .04$, but post-hoc tests, conducted at a Bonferroni-corrected alpha level of .008, did not reveal significant differences between this experiment and others. Cognitive Ability was not different across experiments, $H(2) = 0.29$. However, I found that NFC scores were significantly different across these four measurements, $H(3) = 15.85$, $p = .001$, and post-hoc tests (at an alpha of .008) revealed that the current experiment's participants had significantly lower NFC scores than those in Experiment II, $U = 1613.5$, $z = -2.73$, $p = .006$; no other comparisons involving the current experiment were significant (both zs < 2.20 in magnitude). Considering that these two samples were drawn from the same population – Cardiff University

undergraduates – it is difficult to ascertain why there may have been a difference. Differences in year level and the presence vs. absence of others in the lab may each have played a role; these possibilities will be discussed in greater detail later.

6.2. Results

One participant was removed from analysis for scoring lower than 3 standard deviations from the mean in NFC, and a second was removed for looking at only one judgment per trial; this participant clearly failed to comply with instructions. This left a final *N* of 58 in the Accuracy condition and 58 in the Control condition. Also, when performing analyses involving Median Confidence, I removed from analysis an additional three participants whose Median Confidence was zero, as such participants never interacted with the confidence bar. Further, the distribution of Median Confidence in the Control condition was not normal, and I therefore progressively removed extreme data points in order to achieve a distribution that tested as normal; all three such removed cases were at least 2 standard deviations from the mean and totaled less than 3.5 out of 21. As in previous experiments, if a participant viewed only a single judgment during a trial, I considered it an error in button-pressing; these trials were discarded from analysis. The median number of such errors per participant was 1, and the greatest number of errors made by a single participant was 4.

All dependent measures (Time Taken, Accuracy, Median Confidence, Improvement, AE Focus – Task, and AE Focus – General) were measured in the same way as in Experiments I and II; descriptive statistics for these measures are displayed in Table 6.1. In order to assess any differences across experiments, I compared all conditions that replicated Experiment I (Experiment I, Experiment II – Block 1, and Experiment V – Control condition). I found, first of all, that there were significant differences across experiments for the measure of Time Taken, $H(2) = 19.05$, $p <$

.001, and AE Focus (Task), $H(2) = 9.30$, $p = .01$. I performed Bonferroni-corrected post-hoc tests, at an alpha of .017, in each case. These showed that, first of all, participants completed the task significantly faster in the current experiment (M = 2.65, SD = 0.12; scores logged) than in Experiment I (M = 2.79, SD = 0.13), U = 346.0, z = –4.30, p < .001. Second, participants in the current experiment were significantly more focused on balancing accuracy and efficiency (*Median* = 50.0) than in Experiment II (*Median* = 33.0), U = 179.0, z = –2.95, p = .003; the current experiment's participants returned to an even focus rather than focusing more on accuracy. There were no differences between experiments in Accuracy, Median Confidence, Improvement, or AE Focus (General) (all Hs < 5.91).

Table 6.1. Experiment V: Descriptive statistics for dependent measures, by condition.

	Accuracy		*Control*		*Speed*	
	Mean	SD	Mean	SD	Mean	SD
Time Taken (log)	2.72	0.14	2.65	0.12	2.61	0.11
Accuracy	9.43	3.37	9.36	3.38	9.11	3.81
Median Confidence	11.67	2.81	11.12	2.41	10.50	3.33
Improvement	.00	.26	-.04	.27	-.02	.26
AE Focus - Task	50.00	-	50.00	-	62.00	-
AE Focus - General	42.19	15.15	42.36	16.73	37.98	14.20

Note. AE Focus (Task) was not normally distributed and thus median scores are reported. Low AE Focus scores indicate a focus on accuracy while high scores indicate a focus on efficiency.

The most likely reason that participants performed faster on this task in comparison to an earlier experiment is that, in this experiment, participants were tested in a room full of others, as opposed to alone. The evidence for social facilitation clearly shows that the mere presence of another person, even one who is blindfolded and wearing earphones, tends to speed a participant's performance on simple tasks (see, e.g., Schmitt, Gilovich, Goore, & Joseph, 1985). Although there may seem to be the potential danger that participants have sped to the point that they have reached a common floor in time taken, there was still a great deal of variation in their performances, and therefore room for NFC to explain this variation.

The difference in AE Focus is more puzzling. When I first discovered that participants in Experiment II were focusing more on accuracy than those in Experiment I, I suggested that the university students participating in Experiment II may have felt the setting of the study was more evaluative than the paid participants participating in Experiment I, and therefore decided to focus more on accuracy. However, the students in the current experiment were also university students, taking part in a study within their university building, and yet they focused equally on accuracy and efficiency, as asked. I can only hypothesize that those participants in Experiment II were more likely to have encountered the experimenter (myself) in my role as a Level 1 teaching assistant, as I was during that time teaching students, and this would have emphasized the evaluative nature of the setting. Other than this, the change in year of the students, and the larger lab as opposed to the smaller, there was little difference between the two populations.

In order to determine the suitability of Cognitive Ability as a covariate, I performed a univariate ANOVA with Cognitive Ability as the dependent measure, including condition and NFC (measured)[13] as the independent variables. This analysis revealed no main effects or interactions (all Fs < 1.92), indicating, as in previous experiments, that Cognitive Ability is an appropriate covariate to include. Correlations between Cognitive Ability and primary variables in this study are displayed in Table 6.2. Where significant, these relationships were similar to those in past experiments, with the exception of the relation between Cognitive Ability and Accuracy in the Control condition. In the current experiment, this correlation reached significance; participants higher in Cognitive Ability were also more accurate, $r = .33$,

[13] Again, all analyses involving NFC (measured) were repeated using a median split of NFC (*Median* = 78.5), with similar results.

p = .01. There was no significant effect of Cognitive Ability in any of the main analyses (all Fs < 1.91).

Also, in order to help ascertain whether the manipulation of the instructions was successful, I examined participants' AE Focus during the task. AE Focus (Task) proved to have a bimodal distribution, such that it was significantly non-normal in each of the Control, Accuracy, and Speed conditions, $W(58) = .95$, $p = .02$, $W(58) = .94$, $p = .01$, and $W(54) = .94$, $p = .006$, respectively. It was apparent that participants were reluctant to choose the starting location (50) for their answer. Because of this, I performed a nonparametric ANOVA to examine it, with Condition as a factor. This resulted in a significant main effect, $H(2) = 7.74$, $p = .02$, and post-hoc tests, conducted at a reduced alpha of .017, revealed a significant difference in participants' AE Focus between the Accuracy and Speed groups, $U = 1108.0$, $z = -2.67$, $p = .008$, such that those in the Speed group reported a greater focus on efficiency (*Median* = 62.0) compared to those in the Accuracy group (*Median* = 50.0). I will argue that this evidence, combined with the evidence I discuss below, indicates that participants paid attention to the instructions and attempted to follow them. AE Focus (Task) was also unrelated to NFC in all conditions, $r = .09$, $p = .53$, $r = .10$, $p = .46$, and $r = -.04$, $p = .75$, for Control, Accuracy, and Speed conditions, respectively. Because of its bimodal distribution, I did not include it in further analyses.

Finally, I examined the correlations among the other primary variables (see Table 6.2). Those relationships involving NFC are discussed in the next section, but there were several other relationships of note. Many of these were typical and non-surprising. First, accuracy was marginally and significantly correlated with confidence in the Control and Speed conditions, $r = .24$, $p = .09$, and $r = .35$, $p = .01$, respectively; participants who were more accurate were also more confident.

Table 6.2
Experiment V: Intercorrelations among primary variables.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Ability Measures														
1. Cognitive Ability	–													
2. Need for Cognition	.15 §	–												
Choice Task - Control Condition														
3. Time Taken	.25 §	-.14	–											
4. Accuracy	.33 *	.02	.26 §	–										
5. Median Confidence	.18	.16	.24 §	.24 §	–									
6. Improvement	.03	-.01	.19	.06	-.04	–								
Choice Task - Accuracy Condition														
7. Time Taken	.12	-.07	–	–	–	–	–							
8. Accuracy	-.08	.05	–	–	–	–	.22	–						
9. Median Confidence	-.22	.27 *	–	–	–	–	-.07	-.22	–					
10. Improvement	-.01	.11	–	–	–	–	.13	-.20	.08	–				
Choice Task - Speed Condition														
11. Time Taken	-.05	.07	–	–	–	–	–	–	–	–	–			
12. Accuracy	-.09	.14	–	–	–	–	–	–	–	–	.35 *	–		
13. Median Confidence	-.26 §	.17	–	–	–	–	–	–	–	–	.14	.35 *	–	
14. Improvement	.24 *	-.15	–	–	–	–	–	–	–	–	.07	.31 *	.02	–
AE Focus														
15. AE Focus - General	-.12	.00	-.16	-.41 **	.19	-.11	.12	.07	.13	.00	-.34 *	-.09	-.11	-.22

Note. Outliers were assessed and removed in pairwise correlations by the Jackknifed Mahalanobis distance method (Mahalanobis, 1936; also see text). Time Taken refers to the total time taken to complete all 20 trials. AE Focus refers to the participant's focus on either accuracy or efficiency; a low AE Focus score indicates a focus on accuracy and a high score indicates a focus on efficiency. I do not report between-subject correlations.

§ $p < .10$.
* $p < .05$.
** $p < .01$.

Second, those who reported focusing on accuracy more in general were more accurate in the Control condition, $r = -.41$, $p = .002$, spent more time in the Speed condition, $r = -.34$, $p = .02$. Accuracy in the Speed condition was also significantly related to improvement in that condition, $r = .31$, $p = .03$.

However, in this experiment, unlike in previous, there was a positive relationship between time spent and accuracy, marginally so in the Control condition, $r = .26$, $p = .052$, and significantly so in the Speed condition, $r = .35$, $p = .01$. In Chapter 2 I suggested that a non-significant correlation between time and accuracy is to be expected if participants have a certain threshold of accuracy they aim to achieve; consequently, a positive correlation suggests that perhaps instead they simply spent a fixed amount of time on each trial, or otherwise paid less attention to achieving accuracy. Also different to previous experiments, there was a marginally positive correlation between time spent and median confidence in the Control condition, $r = .24$, $p = .08$; previously confidence was either negatively related to time spent or not related at all.

6.2.1. Main analyses

I first set out to determine the relationship between NFC, Condition, and Time Taken in this experiment, and I thus conducted a 3 (Condition: Control, Accuracy, or Speed) by NFC (measured) ANCOVA, with Time Taken as the dependent measure and Cognitive Ability as a covariate (see Figure 6.1)[14]. It is apparent from the graph that there is a significant main effect of Condition, $F(2, 163) = 11.70$, $p < .001$, $\eta_p^2 =$

[14] Note that, due to violations of assumptions, it was difficult to conduct a reliable ANCOVA including Universe as a factor. When attempted, the only result in this ANCOVA related to Universe was the main effect of Universe, $F(1, 163) = 56.58$, $p < .001$, $\eta_p^2 = .258$, such more time was spent in the weak universe ($M = 2.38$, $SE = 0.01$) than the strong universe ($M = 2.34$, $SE = 0.01$). All interactions involving Universe were nonsignificant (all $Fs < 0.36$). Since Universe was not the main focus of my analyses involving the other dependent measures, it was omitted from these further analyses.

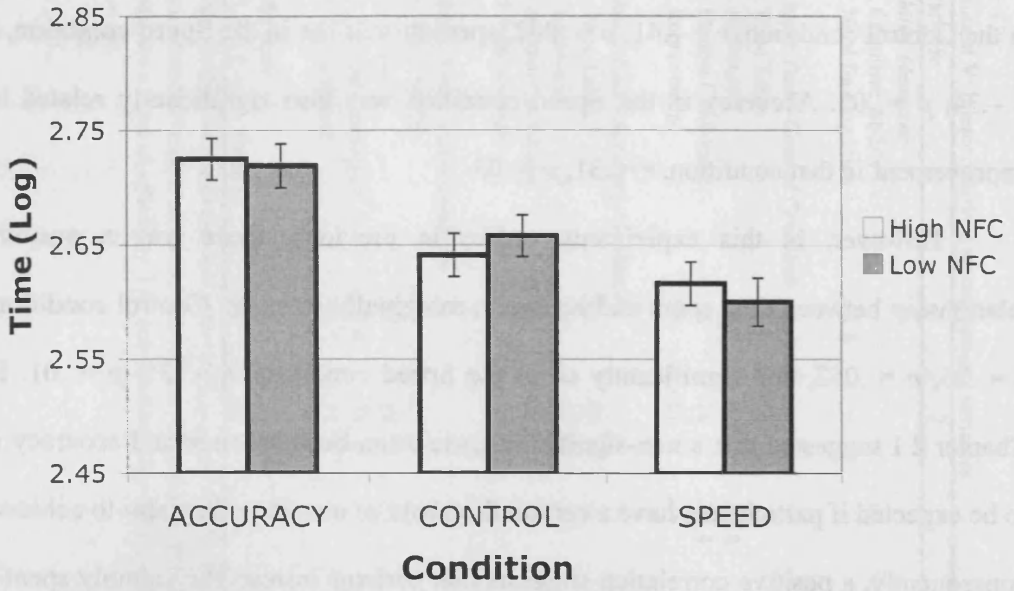


Figure 6.1, Experiment V. Average Time Taken for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

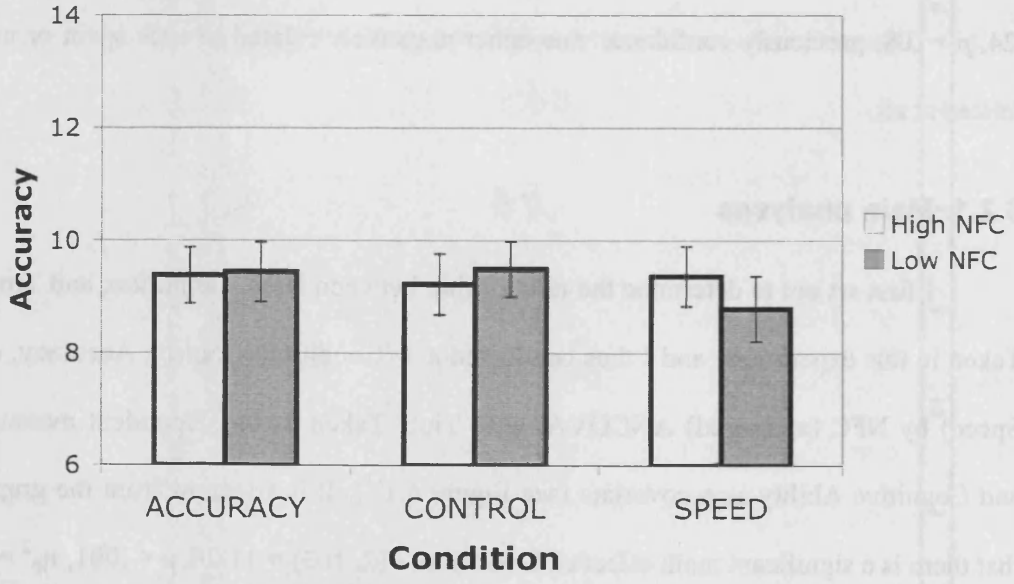


Figure 6.2, Experiment V. Average Accuracy (number correct minus number incorrect) for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

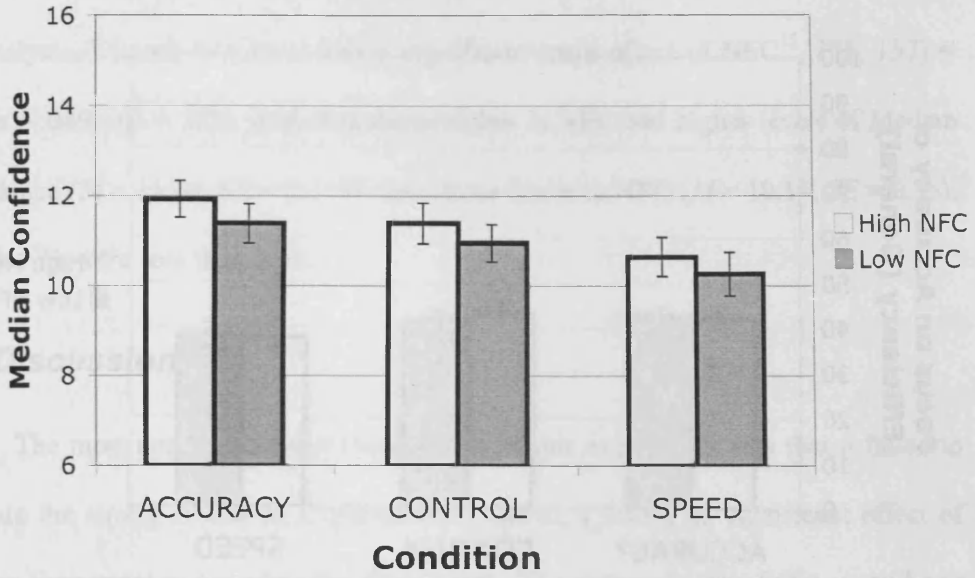


Figure 6.3, Experiment V. Average Median Confidence for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

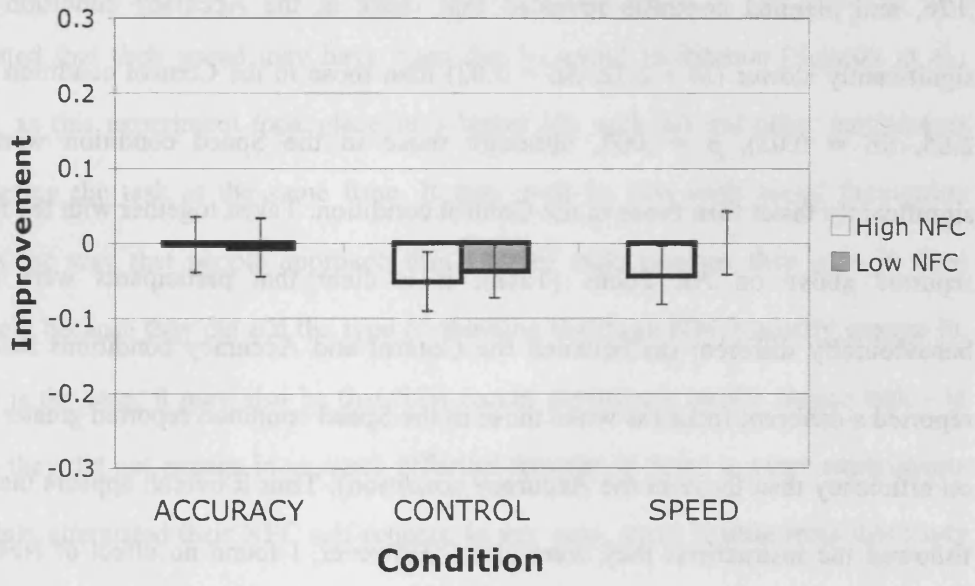


Figure 6.4, Experiment V. Average Improvement score for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

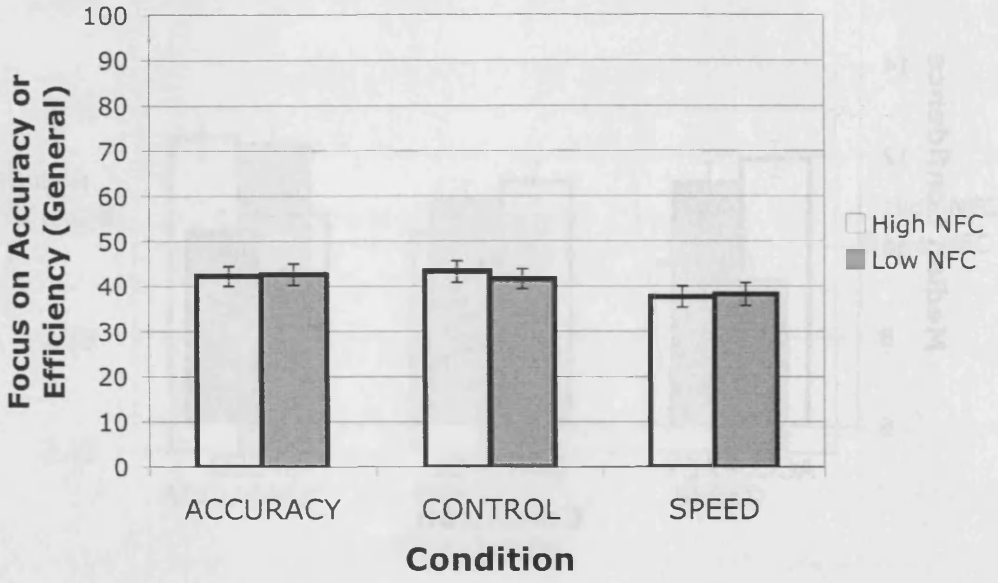


Figure 6.5, Experiment V. Average reported general AE Focus for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. A high AE Focus indicates a focus on efficiency; a low AE Focus indicates a focus on accuracy. Bars are 1 standard error.

.126, and planned contrasts revealed that those in the Accuracy condition were significantly slower (M = 2.72, SE = 0.02) than those in the Control condition (M = 2.65, SE = 0.02), p = .003, although those in the Speed condition were not significantly faster than those in the Control condition. Taken together with the results reported above on AE Focus (Task), it is clear that participants were either behaviourally different (as between the Control and Accuracy conditions here), or reported a different focus (as when those in the Speed condition reported greater focus on efficiency than those in the Accuracy condition). Thus it overall appears that they followed the instructions they were given. However, I found no effect of NFC, nor interaction of Condition with NFC (both Fs < 0.95).

I also conducted identical analyses on the other dependent measures: Accuracy, Median Confidence, Improvement, and AE Focus (General) (see Figures

6.2–6.5). In only one case did I find any significant effects: Median Confidence. For this analysis, I found that there was a significant main effect of NFC[15], $F(1, 157) = 3.93$, $p = .049$, $\eta_p^2 = .024$, such that those higher in NFC had higher levels of Median Confidence ($M = 11.34$, $SE = 0.36$)[16] than those lower in NFC ($M = 10.14$, $SE = 0.36$). All other Fs were less than 2.39.

6.3. Discussion

The most notable and surprising result of this experiment was that it failed to replicate the strong effects of Experiments I and II; I found no significant effect of NFC on time spent to complete the choice task. The reason for this failure may be as simple as a Type II error, chance dictating that I not discover a relationship that is truly present. It may also be related to the differences I discovered between the samples; participants in this experiment both reported significantly lower NFC scores than in Experiment II and were overall faster at completing the choice task. Earlier I suggested that their speed may have been due to social facilitation (Schmitt et al., 1985), as this experiment took place in a larger lab with several other participants completing the task at the same time. It may well be that such social facilitation affects the way that people approach this kind of task; perhaps they gain in time precisely because they cut out the type of thinking that high NFCs usually engage in. If this is the case, it may also be that their recent experience on the choice task – in which they did not engage in as much effortful thought as those in other experiments – slightly attenuated their NFC self-reports. In any case, there is little from this study that I can conclude about the relationship between NFC and time spent.

---

[15] The analysis I did on this dependent variable using a median split of NFC (*Median* = 78.5) had a p-value of .08.

[16] These means, and those for low NFCs, are based on a median split of NFC (*Median* = 78.5).

The only significant result in the main analysis was that NFC was related to participants' confidence, such that participants higher in NFC were more confident, on average, than participants lower in NFC. This result, too, is surprising, given that I found no such relationship in Experiment I or II, and high NFC participants' confidence curves did not differ from low NFCs' in Experiment IV. It may be that I did not have the power to detect the relationship in those experiments, as the effect was small. Or it may be that high NFC participants, when not sampling further to satisfy themselves, compensate by reporting higher confidence for the samples they do accept. This conjecture is in line with the results from Experiment III, wherein participants higher in NFC reported greater confidence in the larger samples they were presented with and less confidence in the smaller samples. It may be that, when not in a situation where they can freely sample (e.g., because to do so requires some extra cognitive justification, as in Experiment III, or because of arousal due to the presence of others, as in the current experiment), high NFC participants reduce their discomfort by convincing themselves that they are more confident than they are. Such management of inconsistent feelings is perfectly in line with cognitive dissonance theory (Cooper, 2007; Festinger & Carlsmith, 1959).

Finally, this experiment provided another opportunity to test for an ability failure of high NFCs. If it is the case that high NFCs are poorer at remembering evidence than low NFCs, then when sampling faster than a control condition, they should have difficulty performing as accurately. Thus, if this were the case, I would have expected to see differences in accuracy between the Accuracy and Control conditions and the Speed condition. However, again, there were no differences. As this was a stronger test than Experiment II's, given that participants were not allowed a full block in which to practice the task before beginning to sample faster, and as I

did not find the effect in either experiment, it is now even less probable that high NFCs have little difficulty remembering faces. At the least I can conclude that these particular high-NFC participants showed no more difficulty than their low NFC peers.

7. Experiment VI: Performance on a small-scale sampling task: The trivia problem

In the previous experiments, I always tested participants in an *external* sampling task; that is, they knew that they were meant to gather (external) evidence and use this to make their judgments. In the current experiment, I test whether their behaviour, in particular the NFC-time relationship, generalizes to an internal sampling task, one in which the participant must consider (or "sample") items from memory in order to determine the answer (Klayman, Soll, Juslin, & Winman, 2006). For example, with the question, "What percentage of people in the U.K. smoke?", the participant may consider acquaintances from the U.K. and use this sample to estimate the population proportion.

Testing this kind of sampling task has several advantages. The first is that, with the access of evidence restricted to memory, there is no possibility of searching for patterns. Also, with the reduced timescale, daydreaming due to boredom, or otherwise thinking about things unrelated to evidence search, is unlikely, as is spending time in order to please the researcher. Thus a propensity to think for longer on this task would be a result of some other reason – a focus on accuracy or a time-spending heuristic. I predicted that I would find such a propensity here.

Accuracy focus, I conjecture, may be related to *adjustment*, the cognitive process of fine-tuning one's opinion until an exact answer can be enumerated. I refer, of course, to the classic phenomenon of anchoring and adjustment (Tversky & Kahneman, 1974). Tversky and Kahneman presented participants with randomly generated numerals; these were followed by questions with numerical answers, such as "What is the percentage of African countries in the United Nations?" The

surprising result was that participants given a higher random number produced higher answers. I suggest that this phenomenon is related to accuracy focus: The more one is focused on obtaining an accurate result, the more one is likely to adjust away from a presented numerical anchor. Although Tversky and Kahneman did not find that offering rewards for accuracy resulted in greater adjustment, an internal drive toward accuracy, as may be found in individuals high in NFC, could produce different results. I thus designed this experiment to replicate Tversky and Kahneman's design. I predicted that, not only would high NFC individuals take longer to answer such trivia questions, they would also adjust further from the presented numerical answers. They may also come closer to the correct answer.

Finally, because including this task alongside the task from Experiment V afforded me the opportunity of dividing participants into groups focusing on accuracy, speed, or both, I did so. I predicted the same pattern of results that I predicted for Experiment V; that is, I predicted that participants in the Accuracy condition would take time similar to high NFCs in the Control condition, and that participants in the Speed condition would take time similar to low NFCs in the Control condition.

7.1. Method

7.1.1. Participants

Participants were the same as those who participated in Experiment V. They were 172 adults (160 female and 12 male), who were recruited from the Cardiff University Participant Panel and received course credit for their time. Their median age was 19. In total, there were 59 participants in the Control condition, 59 in the Accuracy condition, and 54 in the Speed condition.

7.1.2. Design

I presented participants with 28 trivia questions, in two blocks (see Table 7.1). In the first block, they received 12 questions designed to evoke a mental sampling of memory (Sampling Questions), and in the second block, they received 16 questions designed to evoke only a mental search for a potentially known answer (Knowledge Questions). Knowledge questions were included to provide a basis for comparison with Sampling questions, and in this sense acted as control questions. Sampling questions were always of the form, "What percentage of X do Y?", e.g., "What percentage of the world lives in cities?", and Knowledge questions were always of the form, "How many X are in Y?", e.g., "How many seconds are in a minute?" Knowledge questions were additionally divided into (very) easy questions, such as the latter, and hard questions, such as, "How many pints of blood are in the body?" I attempted to match all questions' lengths in characters; the shortest question had 19 characters and the longest had 37, with a median of 23.

All questions had answers in the range of 1-100 (see Table 7.1). I ensured that, for both Sampling and Knowledge questions, half had answers in the "low" range (1-50) and half in the "high" range (51-100). Answers were estimated based on internet sources of at least moderate repute; sources used included institutional web sites (University of Hamburg, Cancer Research U.K., International Dairy Foods Association, Amateur Athletics Foundation of Los Angeles, Office for National Statistics), Wikipedia, Google Public Data, a book web site (*The PC Upgrade Repair Bible*, by Marcia Press and Barry Press), and newspaper and magazine web sites (The Guardian, Computerworld, Next Generation Food). In one case ("What percentage of U.K. people own a computer?"), I estimated the U.K. answer by using the U.S.

Table 7.1. Experiment VI: Questions, Estimated Answers, and Assigned Random Number

Sampling Questions	Answer	# Assigned
What percentage of people in the UK smoke?	21.5	24
What percentage of countries are African?	24.4	17
What percentage of ice cream sold is vanilla?	30.2	39
What percentage of people are left handed?	8.5	77
What percentage of waste is food waste?	19	57
What percentage of land surface is forested?	30	72
What percentage of children play a sport?	96	12
What percentage of UK people are Christian?	71.8	50
What percentage of people have brown eyes?	92	10
What percentage of the world lives in cities?	59	55
What percentage of UK people own a computer?	65	86
What percentage of computers run Windows?	89.6	94
Knowledge Questions		
How many months long is a pregnancy?	9	24
How many days are in September?	30	8
How many eggs are in a dozen?	12	73
How many letters are in the alphabet?	26	81
How many seconds are in a minute?	60	36
How many degrees are in a right angle?	90	25
How many weeks are in a year?	52	63
How many years are in a century?	100	59
How many pints of blood are in the body?	10	36
How many chromosomes do humans have?	46	11
How many people are on a rugby team?	15	63
How many countries are in the EU?	27	57
How many millions of people are in the UK?	62	37
How many years does the average UK person live?	79.9	16
How many states are in the US?	50	96
How many years old is the Queen?	83	63
Median (Sampling Questions)	44.6	52.5
Median (Knowledge Questions)	48	47

Note. Assigned random number denotes the random number that was the final number participants viewed onscreen. See text for the derivation of estimated answers and the design of number allocation.

answer. In three cases ("What percentage of children play a sport?", "What percentage of ice cream sold is vanilla?", "What percentage of the world lives in cities?"), I estimated world or "general" figures by using country figures (U.S., U.S., and all countries averaged, respectively). Finally, in one case ("What percentage of people have brown eyes?"), I estimated an answer based on Wikipedia's U.S. answer, adjusted to account for countries of primarily brown-eyed peoples.

Each question was paired with a series of five numbers, each of which was between 1 and 100. The first four numbers were all randomly determined with no restrictions, while the last number was randomly determined to either be "low" (1-50) or "high" (51-100). Questions were thus paired with numbers to fit a 2 (Answer: High or Low) by 2 (Random Number: High or Low) design in the case of Sampling questions, and a 2 (Answer: High or Low) by 2 (Random Number: High or Low) by 2 (Difficulty: Easy or Hard) design in the case of Knowledge questions. All such numbers were predetermined and thus the same for all participants. The purpose of this design was to ensure that all participants saw numbers from the complete range and that the questions' answers also covered the complete range.

7.1.3. Procedure

Participants first completed the task comprising Experiment V, during which they were assigned a condition: Control, Accuracy, or Speed (see Experiment V, Method). These conditions were preserved for this experiment, such that a participant, e.g., in the Control condition in Experiment V, received instructions to focus on accuracy and efficiency equally during this task as well, while participants in the Accuracy and Speed conditions received instructions focusing on either accuracy or efficiency, respectively (see Appendix F for the full instructions). Participants were always tested in a large lab in groups of 10-15, with the exception of one group of 4.

Before each block, participants were informed that they would be answering trivia questions and given an example question of the same format as those in the block. They were also informed that each question would be preceded by a series of random numbers, with one number remaining on screen, and that all answers would be in the 1-100 range. They then received detailed information on how to fill in their answers.

Each trial began with the presentation of the five random numbers; each number was displayed for 1000 milliseconds, with the exception of the final number, which remained onscreen. 1000 milliseconds after the final number (X) appeared, the question appeared along with two options: "Above X?" or "Below X?" E.g., the question "What percentage of children play a sport?" was paired with the number 12 and thus accompanied by "Above 12%?" or "Below 12%?"; this was in order to replicate Tversky and Kahneman's (1974) method. Once the participant selected an answer, a new question appeared, asking, "Actual Percentage?" or "Actual Number?" depending on block; this was accompanied by a blank box in which to fill the answer. I timed the participant in seconds from the moment the questions appeared to the time the participant clicked the "Next" button to move to the next trial.

When participants had completed all trials, they then completed questionnaires (including a question regarding their focus on accuracy or efficiency for this task), and, following this, they were asked to read from an online novel (*Pride and Prejudice*, by Jane Austen) while waiting for their fellows to finish. They returned between one and seven days later to complete the Raven matrices; this was done in groups of 1 to 3.

7.1.4. Materials

The Raven matrices, NFC and SDR questionnaires used in this experiment were identical to those in Experiments I-V. Because these participants were the same as those who completed Experiment V, scores are identical to those reported for Experiment V. Mean scores were –1.51 (*SD* = 7.70) for the Raven matrices, 78.95 (*SD* = 14.25) for the NFC measure, and 16.61 (*SD* = 4.82) for the SDR measure. Cronbach's α for these measures were .76, .93, and .66, respectively. See Experiment V (Method) for a discussion of how these scores differed from those in other experiments.

7.2. Results

I removed one participant from analysis for scoring lower than 3 standard deviations from the mean in NFC, and I removed a second participant who did not record any numerical answers; this participant clearly failed to comply with instructions. This left final *N*s of 58 in both the Accuracy and Control conditions.

The dependent measures I examined in this experiment were the reaction times to complete the questions (Knowledge RT, Sampling RT), the difference in reported answer from the provided random number (Knowledge Difference, Sampling Difference), and the difference in reported answer from the estimated correct answer (Knowledge Accuracy, Sampling Accuracy). In each case, I took the median of the participant's responses in order to create that participant's contribution to the analysis. For the Knowledge measures, in each case I examined these split into Easy and Hard conditions; in the case of the Easy condition, Knowledge Difference and Knowledge Accuracy showed very little variance, as nearly all participants knew the correct answer; as such, I did not analyze them further. Two of the remaining measures

(Knowledge RT and Knowledge Accuracy) proved to have distributions that were significantly different from normal in nearly all conditions, and I therefore report log-transformed values. One measure, Sampling Difference, displayed significantly heterogeneous variance across conditions, and I also log-transformed this measure to compensate. In several cases (Knowledge RT – Easy, Knowledge RT – Hard, Sampling RT, Sampling Difference, Sampling Accuracy), I was additionally forced to remove some outliers in order to achieve a normal distribution. I did so by progressively removing extreme data points until the distribution tested as normal; each outlier removed in this way was at least 2 standard deviations from the mean, and the largest number of outliers I removed from any one condition was 3. Unfortunately, it was not possible to log-transform all dependent variables for consistency's sake, as in some cases this resulted in the transformation of a normally distributed variable to a non-normally distributed one. Thus, for ease of comparison, in Table 7.2, I report the descriptive statistics in untransformed state. Where later analyses are based on log-transformed values, I have used the statistics from the log-transformed distributions, but reverted them.

Table 7.2. Experiment VI: Descriptive statistics for dependent measures, by condition.

	Accuracy		*Control*		*Speed*	
	Mean	SD	Mean	SD	Mean	SD
Knowledge RT - Easy	6.61	1.20	6.61	1.20	6.31	1.20
Knowledge RT - Hard	8.32	1.35	8.32	1.35	7.24	1.23
Sampling RT	10.23	2.41	10.37	2.64	9.61	2.31
Knowledge Difference - Hard	29.29	8.33	29.16	8.74	28.01	5.92
Sampling Difference	14.79	1.51	14.79	1.51	14.13	1.35
Knowledge Accuracy - Hard	5.50	1.86	14.79	1.51	14.13	1.35
Sampling Accuracy	19.82	4.89	19.42	4.86	17.78	4.65
AE Focus - Q Task	38.00	-	47.00	-	39.00	-

Note. Knowledge RT – Easy, Knowledge RT – Hard, Sampling Difference, and Knowledge Accuracy – Hard were log-transformed initially for future analyses. The scores reported here are those of the log-transformed distributions, reverted to regular scoring for ease of comparison. AE Focus (Q Task) was not normally distributed and so median scores are reported; high scores here indicate a focus on efficiency while low scores indicate a focus on accuracy. RT scores are reported in seconds.

One final dependent measure I examined was AE Focus (Q Task), the participants' reported focus on accuracy or efficiency during the question-answering task. As in the previous experiment, this distribution was significantly non-normal in two conditions, $W(58) = .94$, $p = .01$ (Control), $W(54) = .95$, $p = .03$ (Speed); it was clearly bimodal, for the same reason as in Experiment V (participants' refusal to respond near the starting point of the scale, at 50). I thus chose to examine it only as a manipulation check, using a nonparametric ANOVA with AE Focus as the dependent measure and Condition as the factor. However, the effect of Condition was only marginal, $H(2) = 4.98$, $p = .08$, possibly due to the reduced power of the nonparametric test.

In order to determine the suitability of Cognitive Ability as a covariate, I referred to the univariate ANOVA I performed for Experiment V, with Cognitive Ability as the dependent measure, Condition and NFC (measured)[17] as the independent variables. As a reminder, this analysis revealed no main effects or interactions, meaning that Cognitive Ability is an appropriate covariate to include. Cognitive Ability did not contribute significantly to any of the analyses below (all Fs < 2.31).

7.2.1. Main analyses

First, I examined the relationship between NFC and my measures of time: Knowledge RT and Sampling RT (see Figures 7.1 and 7.2). For the former, I performed a 3 (Condition: Control, Accuracy, or Speed) by 2 (Difficulty: Easy or Hard) by NFC (measured) ANCOVA, with Knowledge RT as the dependent variable

[17] Again, all analyses involving NFC (measured) were repeated using a median split of NFC (*Median* = 78.5), with similar results.

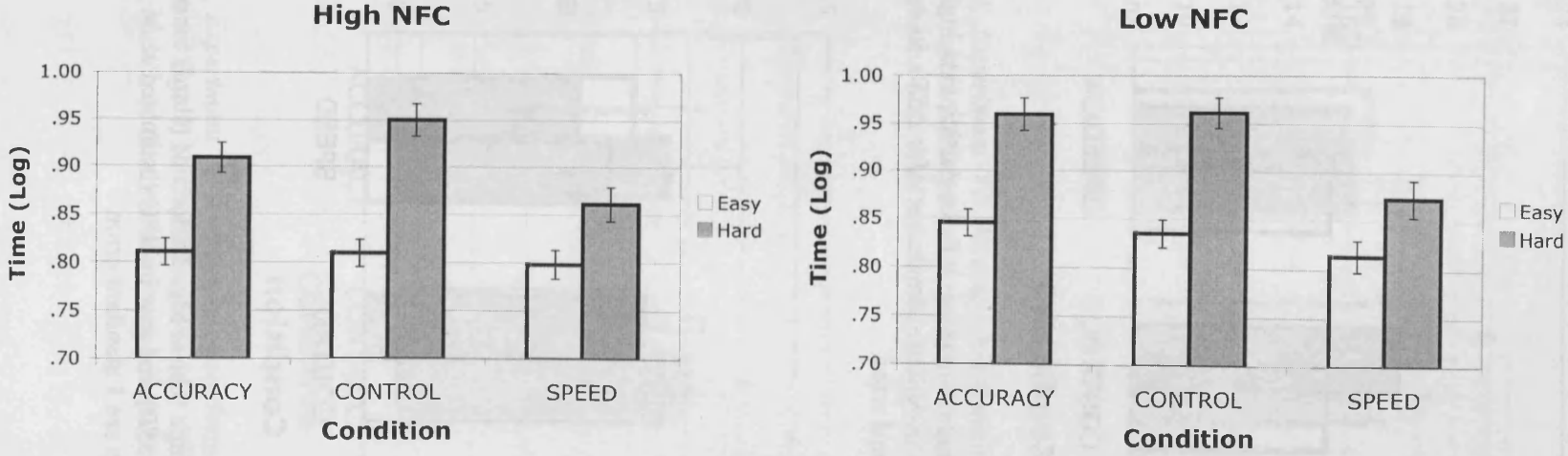


Figure 7.1, Experiment VI. Average log-transformed Knowledge RT for participants high in NFC (left; estimated at $M + .5SD$) and low in NFC (right; estimated at $M - .5SD$), in each of the three conditions and for both Easy and Hard question types. Bars are 1 standard error.

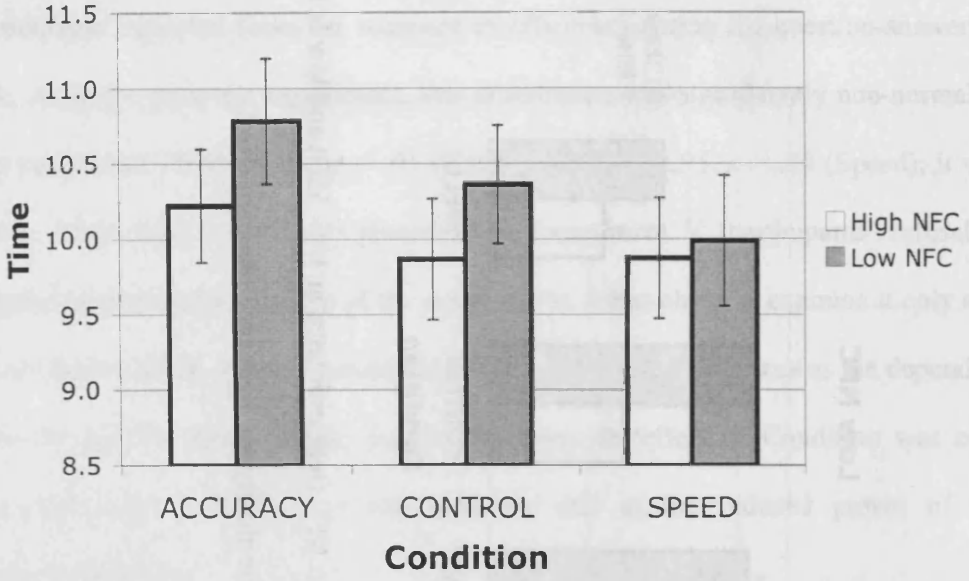


Figure 7.2, Experiment VI. Average Sampling RT for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

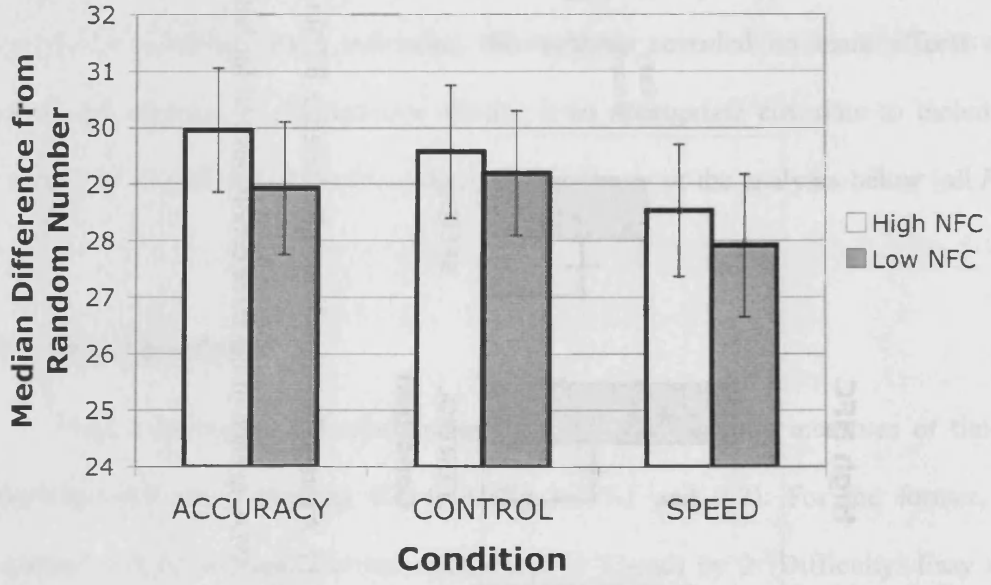


Figure 7.3, Experiment VI. Average Knowledge Difference (Hard) for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

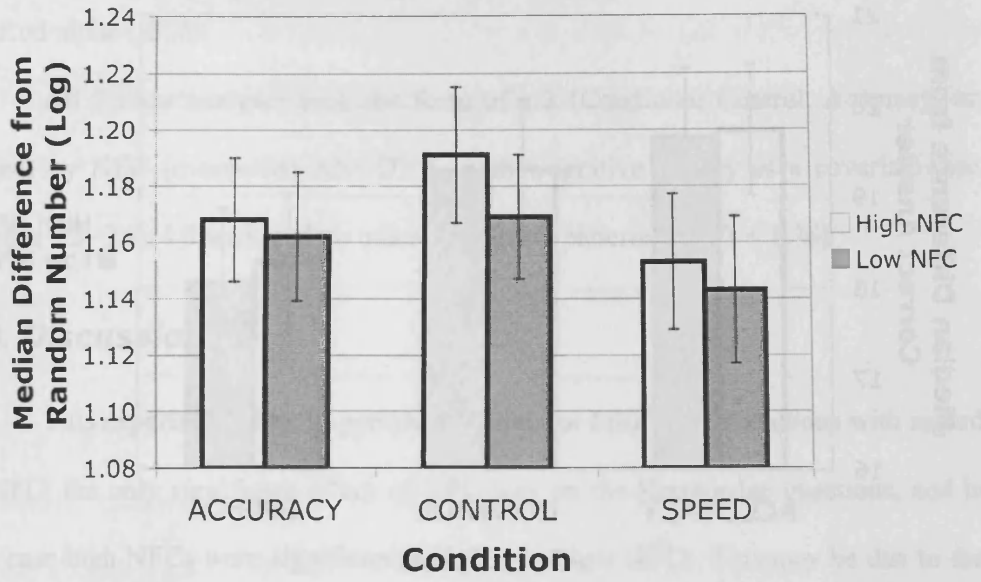


Figure 7.4, Experiment VI. Average log-transformed Sampling Difference for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

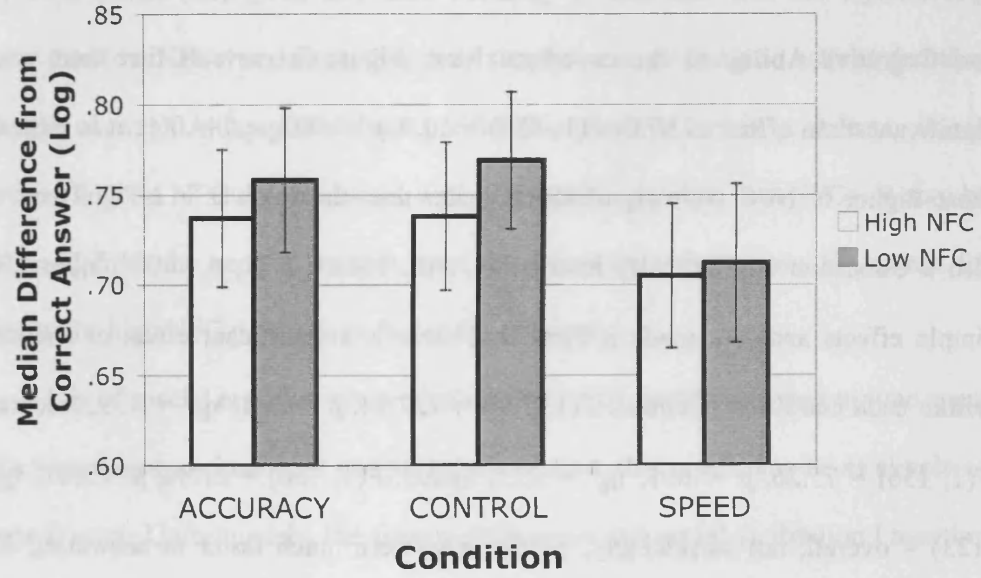


Figure 7.5, Experiment VI. Average log-transformed Knowledge Accuracy (Hard) for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

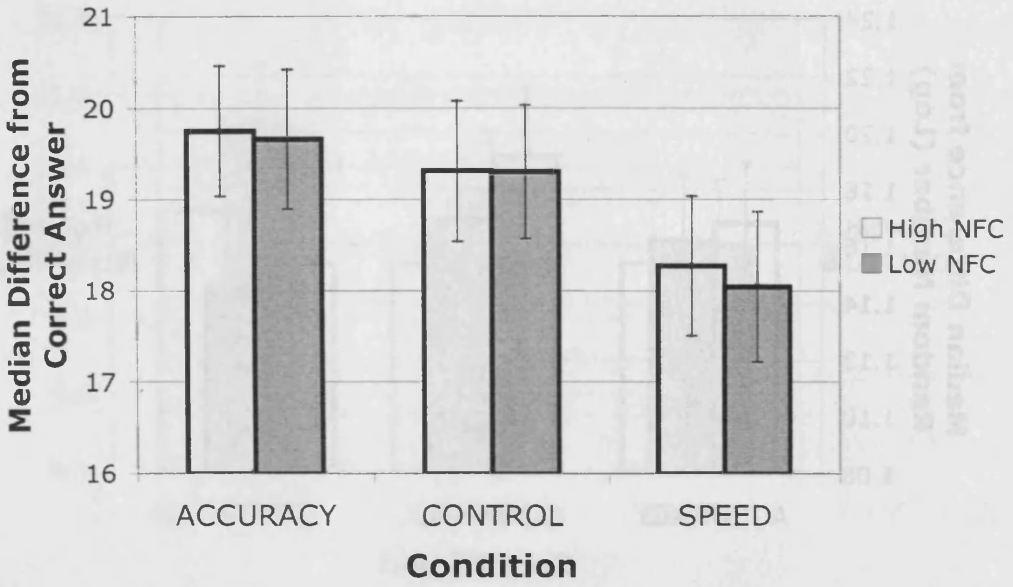


Figure 7.6, Experiment VI. Average Sampling Accuracy for participants high in NFC (estimated at $M + .5SD$) and low in NFC (estimated at $M - .5SD$), in each of the three conditions. Bars are 1 standard error.

and Cognitive Ability as the covariate. First, Figure 7.1 reveals that there was a significant main effect of NFC, $F(1, 156) = 10.0$, $p = .002$, $\eta_p^2 = .060$; it is clear that those higher in NFC were significantly faster than those lower in NFC. There was also a Condition by Difficulty interaction, $F(2, 156) = 8.32$, $p < .001$, $\eta_p^2 = .096$; simple effects analyses made it clear that there is a significant effect of Difficulty within each condition (Control: $F(1, 156) = 122.14$, $p < .001$, $\eta_p^2 = .439$, Accuracy: $F(1, 156) = 75.26$, $p < .001$, $\eta_p^2 = .325$, Speed: $F(1, 156) = 21.84$, $p < .001$, $\eta_p^2 = .123$) – overall, not surprisingly, participants were much faster in answering Easy questions compared to Hard. Further simple effects showed that there was also an effect of Condition within the Hard answers, $F(2, 156) = 8.84$, $p < .001$, $\eta_p^2 = .102$. Thus, when participants were answering Hard questions, they were significantly faster in the Speed condition compared to the Control condition, $t(107) = 3.86$, $p < .001$, or

Accuracy condition, $t(107) = 2.69$, $p = .008$, both significant at a Bonferroni-corrected reduced alpha (.017).

All further analyses took the form of a 3 (Condition: Control, Accuracy, or Speed) by NFC (measured) ANCOVA, with Cognitive Ability as a covariate (see Figures 7.3–7.6). I discovered no other significant patterns (all Fs < 1.76).

7.3. Discussion

This experiment, like Experiment V, did not fulfill my predictions with regard to NFC; the only significant effect of NFC was on the Knowledge questions, and in this case high NFCs were significantly faster than low NFCs. This may be due to the tendency of high NFCs to accumulate information (Inman et al., 1990); their general knowledge may have been slightly superior, and therefore their memory search may have been faster. (Or, given that their accuracy at this task was not significantly greater than low NFCs' accuracy, they may simply be more likely to believe that they know the answer.) Although I did not predict this result, it certainly, at least, proves that those high in NFC do not always deliberate for longer than low NFCs. It may be that, freed from the need to sample more to please an experimenter, or search for patterns, the high NFC person is in fact more efficient.

It is of special note that my prediction that NFC would be related to time spent in the Sampling questions went unfulfilled. As before, this result may be as simple as a Type II error. Unfortunately, the sample differences and social facilitation I mention in discussing Experiment V (see Experiment V – Discussion) may also have been causes. It is impossible to know if this result reflects the truth, fails to detect a difference by chance, or is the result of a dampening in spending time due to arousal.

Finally, once again, I have searched for a potential focus on accuracy, and found none. This result adds to the mixture of results I have found so far; many have

been null (Experiments III, IV, and now VI), and some have been significant, as with the significant relationship of NFC to self-reported accuracy focus in Experiments I and II, and the significantly extra time high NFCs were willing to spend on weak-universe trials in Experiment II.

8. Overview and General Discussion

The purpose of my investigations was to discover whether Need for Cognition (NFC), a thinking style now established as having a unique role in eliciting normative responding on many reasoning tasks, might ever be linked to non-normative behaviour. I thus set out to establish whether, first of all, it is related to the expenditure of time on an information-gathering task, and if so, what reason participants might have for spending this time (as some reasons indicate normative responding and others do not). In Experiments I and II, I established that NFC is indeed uniquely related to time spent (independent of fluid intelligence), such that participants high in NFC (high NFCs) spend more time gathering information than those low in NFC (low NFCs), and importantly, they gain no statistically significant benefit in accuracy in return. Therefore, in a crucial sense, they are failing somehow at this task: They are either failing in their ability a) to remember the evidence, b) to judge their own capability to improve at the task within the time allotted, c) to estimate how much time has passed (see Baugh & Mason, 1986), or they are failing to apply a normative strategy. In my discussion of Experiment I, I then laid out seven potential hypotheses for the reason participants might behave in this manner:

1) *Ability Failure.* High NFCs might simply need more time to reach equivalent accuracy to low NFCs.
2) *Accuracy Focus.* High NFCs might be searching for longer in order to improve their accuracy; i.e., they highly value accuracy.
3) *Improvement.* High NFCs may be attempting to improve at the task, such that their performance on later trials will be better than on earlier trials.

4) *Pattern-searching.* High NFCs may be searching for patterns in the data they are gathering.

5) *Boredom.* High NFCs are bored and daydreaming during the task.

6) *Social Desirability.* High NFCs have a greater desire to appear in a positive light than low NFCs and therefore spend time because they believe this will be pleasing to the researcher.

7) *Time Heuristic.* High NFCs possess a heuristic that suggests to them that they must persist; this is equivalent to an inability to disengage from the task.

Ability Failure and Improvement, I argue, represent normative or potentially normative behaviour, where normative behaviour for this task is defined as adopting the strategy that has the best probability of optimising both accuracy and efficiency (as valued by the experimenter). In the case of Ability Failure, the high NFCs assume the same (presumed normative) strategy as the low NFCs, but simply lack the ability to carry it out as quickly. The Improvement hypothesis suggests normativity so long as the benefits one gains on later trials outweigh the cost of learning to improve on the earlier trials. If high NFCs are adopting this strategy, their failure is in their ability to judge how much time (and ability) they have to improve.

The other five hypotheses, meanwhile, represent non-normative performance. High NFCs spend more time than low NFCs, but do not achieve significantly better accuracy; if this is due to a focus on accuracy, it is leading them away from the optimal accuracy/speed ratio (and therefore away from normativity). Pattern-searching, while it may be useful on some tasks, is pointless in a task involving random data selection. Bored daydreaming is certainly not useful for fulfilling the goals of the task, nor is spending extra time solely to appear in a positive light.

Finally, spending time because of a time heuristic or inability to disengage also does not contribute to completing the task in both an accurate and efficient manner; it only adds time without necessarily gaining in accuracy (or, if gains are made, they may not be of the right amount). Determining why high NFCs spend this time, then, should tell us whether their behaviour is normative.

8.1. Summary of evidence

The best possible ways to test the above hypotheses are as follows: 1) to measure precisely, if possible, these intra-psychic variables, and search for a potential mediator role for them, and 2) to manipulate the variables – or the environment with which the variables interact – and search for a change in behaviour. My experiment plan followed these strategies, with a major focus on Ability Failure, Improvement, and Accuracy Focus. I focused on the former two hypotheses because they were the only two suggesting normativity, and Accuracy Focus because I deemed it the most likely hypothesis. As a secondary focus, I also measured and searched for a mediating relationship involving social desirability (SDR), as there exist validated measures for this. Below, I summarise the manipulations and measurements I did to test each hypothesis, and the results of each.

8.1.1. Ability failure

In this case, I hypothesized that perhaps high NFCs were using the extra time at the task because they required more time to reach the same level of accuracy as the low NFCs. The simplest manipulation available to affect one's ability to complete the task well is to make the task either easier or more difficult. However, making the task easier was not an effective way to test this hypothesis. First, if I had still found an NFC-time relationship for an easier version of the task, I could not be certain that my

manipulation was strong enough. If I had found no NFC-time relationship, I could likewise not be certain that changing the task substantially (e.g., by having faces remain on screen instead of disappearing) did not have some other effect (e.g., making the task less boring, or extraneous pattern-searching easier). I therefore chose to make the task more difficult. I did this, in Experiment II, by presenting one group of participants with a set of instructions requesting them to complete the second trial block more quickly. I replicated this method in Experiment V in a between-subjects manner: There was one condition in which participants were asked to focus on efficiency (as opposed to accuracy or both accuracy and efficiency). In both cases, despite evidence that participants followed the instructions and completed the task faster, I found no evidence that participants' accuracy was negatively affected when they spent less time, which I would have expected if they required more time to reach the same level of accuracy as low NFCs. In Experiment II, the accuracy of high NFCs even improved slightly during the speeded block. Thus, I found no evidence for ability failure. The evidence is, rather, more consistent with the other hypotheses: Perhaps the high NFCs chose to value accuracy less, to stop thinking about other things, to please the experimenter by going faster rather than slower, or moderated their heuristic to spend time. Although ultimately a null result, and therefore possibly a Type II error, it is notable at least that I searched twice and both times failed to find the effect. Combined with the argument that high NFCs, who report enjoying complex problems, are generally not predicted to perform poorly at simple tasks, this hypothesis certainly does not receive much support.

8.1.2. Accuracy focus

In the case of Accuracy Focus, I predicted that high NFCs would value accuracy more than low NFCs and therefore pay the time cost needed to attain it. I

attempted both of my outlined tactics here: I tried to measure accuracy focus precisely and I directly manipulated it. I began by giving participants, in Experiments I and II, a sliding scale on which to measure their focus on either accuracy or efficiency (AE Focus), on either the task itself or in general. As this measure is not validated, I used it only as a guidepost, to indicate the potential for conducting further experiments involving accuracy focus. In Experiment I, I found that AE Focus (General) was significantly related to NFC, even with cognitive ability partialed out, and in Experiment II, I found a main effect of NFC on AE Focus (Task), with cognitive ability's variance removed. Although I found no significant mediation by AE Focus in either case, I still chose to examine accuracy focus further.

I first found that, in Experiment II, participants high in NFC were taking longer to examine weak evidence trials than strong ones, while those low in NFC were spending roughly equal time on the trials, regardless of evidence strength. This is evidence that high NFCs are truly more focused on accuracy than low NFCs; they cared enough, when evidence was weak, to devote more time to the trial. I.e., they cared about accuracy more than their low-NFC peers.

I next attempted a true measure of threshold in Experiment III, in which I presented participants with samples of fixed size and varying evidence strength and asked them on which trials they would prefer to gather more evidence. A focus on accuracy would have been indicated by a general preference for stronger evidence; i.e., participants desiring more accuracy should, on average, be happy to make a choice only when evidence is strong. The results of this experiment trended in the predicted direction, showing that high NFCs preferred to make choices, on average, when evidence strength was higher, but this trend was small and non-significant. Importantly, these participants had just recently completed Experiment II, for which

there was a definite NFC-time effect – and during which the high NFCs showed a preference to spend more time when evidence was weak – and yet they did not require significantly stronger evidence in this task. However, in order to ensure participants would sometimes answer that they wished to gather more data, I had asked them to focus on accuracy in this experiment, and this may have washed out potential differences in accuracy focus. Thus, the results of this experiment are mixed, though there is no evidence in favour of accuracy focus on the part of high NFCs.

Next, I attempted to look for accuracy focus in the way that people's confidence grew during the task (Experiment IV). In particular, a focus on accuracy should be accompanied by a dampening of confidence, in relation to a fixed endpoint or threshold of choice, which was provided in the experiment. This should result in a shallower and longer lasting curve. Again, in this experiment I found a slight trend in the predicted direction, such that high NFCs displayed lower confidence at all levels, but it was small and non-significant. I noted that participants in this experiment might be under significant cognitive load due to attempting to report confidence simultaneously with the task, which may have affected the results. However, again, there was no significant evidence in favour of Accuracy Focus.

Lastly, I attempted to manipulate accuracy focus. In Experiments V and VI, I employed a between-subjects design with three conditions: one that requested participants focus on accuracy (Accuracy), one that requested a focus on efficiency (Speed), and one that placed equal emphasis on accuracy and efficiency (Control). Although my manipulation checks indicated that people indeed employed these different foci (and this translated into significantly more time spent in the Accuracy condition compared to the Control and Speed conditions), I found no significant effect of NFC in either of these experiments; in particular, Experiment V failed to replicate

the main finding of the NFC-time relationship from Experiments I and II. I suggested that the reason for these null findings, if they were not due to chance, may have been that participants were tested in a larger lab, where social facilitation may have influenced them. I.e., participants may have felt arousal due to the others present, which caused them to speed up on the simple tasks they were presented with – wiping out any differences between high and low NFCs.

In sum, I found mixed evidence. However, importantly, although I failed to find evidence in favour of Accuracy Focus several times, these are null results. It is of key importance that, in fact, I did find significant evidence in favour of this hypothesis in Experiment II, where high NFCs were more sensitive to evidence strength than low NFCs. This evidence supports the Accuracy Focus hypothesis, indicating that high NFCs display some level of non-normativity because of their desire for stronger evidence.

8.1.3. Improvement

I hypothesized that perhaps those high in NFC were using their extra time to learn how to perform the task more effectively, such that on later trials they might be more likely to discover the correct answer. I monitored this type of improvement throughout Experiments I, II, IV, and V, the experiments in which participants performed the identical task of gathering information. The only experiment for which I found a difference in improvement score was Experiment II, in which high NFCs attained a higher score than low NFCs. However, the score of high NFCs here was near zero, while low NFCs had an average score that was slightly negative, indicating only that high NFCs were able to maintain performance at the task, while low NFCs slipped slightly. Thus I found no evidence that high NFCs were spending extra time in order to improve at the task.

8.1.4. Pattern-searching

The Pattern-searching hypothesis suggests that high NFC participants are expending more time because they are searching for patterns in the data. This hypothesis is difficult to test directly, as there is no good way to measure it during the task, and it was not the main focus of this thesis. However, I note that it is consistent with the pattern of data I have observed thus far. It may have been that high NFCs spent longer in Experiments I and II (Block 1) due to searching for patterns in the data, but they gave this up in Experiments II (Block 2), IV, V, and VI, when they were either requested to decide faster, put under cognitive load, or subjected to social facilitation. In Experiment III, they were not timed, and any pattern-searching would not have likely affected their actual choice behaviour.

8.1.5. Boredom

The Boredom hypothesis suggests that high NFCs are dividing their attention, such that some of it is focused on the task but the rest is devoted to thinking about task-unrelated things. This hypothesis is also difficult to test; while a measure of boredom exists (Farmer & Sundberg, 1986), it is a stable trait measurement and not a measurement of boredom during a task. It is likewise difficult to manipulate boredom reliably; one could ask participants to be more or less engaged with the task, but it's uncertain what result this would have. However, I note that high NFCs in Experiment II were able to maintain their accuracy performance better than low NFCs, based on their improvement scores. This suggests engagement with the task was consistent, which I would not expect if participants were bored. True boredom suggests that the task is not important enough to devote one's full attention to it, and also suggests that it is not important to maintain engagement with it; attention should gradually slip

further away, especially if mental fatigue sets in. It is possible that high NFCs are less susceptible to mental fatigue, but even without it a careful maintenance of performance is at odds with boredom.

It may still be that high NFCs are simply experts at dividing their attention, such that they are continuously engaging in thoughts unrelated to the task while maintaining performance on the task. However, the maintenance implies a fixed level of importance assigned to the task. We might then reject Boredom, and give this hypothesis a new name: Unrelated Thoughts. The Unrelated Thoughts hypothesis is consistent with the evidence presented for the exact reasons the Pattern-searching (or, we could call it, Related Thoughts) hypothesis is.

8.1.6. Social desirability

I hypothesized that, especially since NFC is typically weakly related to social desirability (SDR) (Cacioppo et al., 1996), one reason for the NFC-time relationship is that participants feel they are expected to expend effort (and therefore time), and do so in part because of this expectation. Throughout my experiments, I measured participants' social desirability, i.e., their tendency to want to present themselves positively to the researcher, and, where I found the NFC-time relationship, I searched for mediation by SDR. I found that indeed, in Experiment II, there was partial mediation of the relationship by SDR. Thus it's highly likely that high NFCs were devoting time to impression management.

Although this behaviour may lead to a slight increase in accuracy (one I did not detect in any experiment, but may be present in general), I argue that it is still non-normative. Normativity, as I have defined it, depends on the experimenter's defined task goals: in this case, performing both accurately and efficiently. However, this finding suggests that there is at least one reason for spending time – social

desirability – that is not related to either. One could imagine that a participant divines that the best way to please the researcher is to optimise for both speed and accuracy – but this is clearly not the case, as the high NFCs fail to do so, getting poorer accuracy/speed ratios than low NFCs. Rather, it appears that high NFCs translate their need for impression management directly into time expenditure alone. In short, because the high NFCs are optimising for a goal outside of accuracy or speed, which worsens their accuracy/speed ratio, the behaviour cannot be normative. Thus again I have some evidence of non-normative behaviour by high NFCs.

As an aside, I should note that, because I used items only from the impression management subscale of SDR, I have been discussing SDR as such: a measure of self-presentation to others, rather than self-deception (which is covered in the other subscale, Paulhus, 1991). Measures of impression formation are designed to include overt behaviours (e.g., lying to others) and therefore "any distortion is presumably a conscious lie" (Paulhus, 1991, p. 37). However, consider the items of the scale: "I always obey rules even if I probably won't get caught," "I like everyone I meet," "I sometimes tell lies if I have to" (reverse scored), "There are times I have taken advantage of someone" (reverse scored), and "I have said something bad about a friend behind his or her back" (reverse scored). Paulhus claims that anyone saying, for example, that he does not take advantage of people, must be lying. But, of course, there may in fact be a small proportion of the population that does not lie or take advantage of others, obeys rules even when unlikely to be caught, doesn't gossip, and/or likes everyone. It should certainly be considered that high NFCs fall into this category. In this case, rather than spending time in order to create a positive impression of himself, the high NFC may be spending this time because he truly believes it is better to expend effort/time on problems (as he may truly believe it is

good to follow rules), or because he wants the experimenter to be pleased out of a sense of compassion. The behaviour is still non-normative, for the reasons argued above, but these two possible conceptualisations should be disentangled with future research.

8.1.7. Time heuristic

The Time Heuristic hypothesis states that the high NFCs possess a heuristic that suggests to them to persist at the task for a long time, perhaps because this strategy is successful for reasoning problems, or other problems that they may be more used to. This hypothesis is perhaps the most difficult of all to test directly, although I will discuss later some possibilities for such testing in future research. I did, however, attempt to test it indirectly, by employing a situation where most of the other reasons for spending time are unlikely to be present (Experiment VI). I did so by presenting participants with a task that involved "sampling" from memory: the trivia problem. Because the items sampled are from memory, there is no point in searching for patterns in them, and due to the short time-scale of the problem, the participant is unlikely to spend time daydreaming, lengthen time spent to please the researcher, or attempt to improve. Thus, I argue, any time spent is likely due either to an accuracy focus or a time heuristic; if Accuracy Focus could later be eliminated, this would provide some evidence for Time Heuristic. However, I found no evidence of the NFC-time relationship in this experiment, and I am uncertain to what degree this may have been due to chance or social facilitation (as discussed earlier).

8.1.8. Overall

Overall, then, I have found strong evidence that high NFCs perform poorly on this task, spending more time than low NFCs but gaining no significant accuracy

advantage. As to why, I have found evidence to support two hypotheses (Accuracy Focus, Social Desirability) and to reject another (Boredom). Thus, the picture I can paint of the high NFC is as follows. This person is motivated, not only for complex tasks and difficult reasoning problems, but also for the simple information-gathering task presented in my experiments, and she takes care to try to obtain correct results even when it costs her time. She is not bored but able to maintain a high level of performance even after many trials. She also, however, expends more time than her low-NFC fellows, at least in part because she has an impression of herself to maintain: one of an effortful thinker who spends time thoroughly when asked – presumably, often, in order to discover the normative solution to a problem. Ironically, in maintaining this impression, for this problem she strays from the normative strategy that would result if she were expending her mental effort solely on the task itself. Further, her focus on accuracy clearly goes beyond what the experimenter desires, and this too means her strategy is non-normative.

8.2. Discussion

I originally set out to discover whether a particular thinking style – NFC – might ever be uniquely related to non-normative performance. I have found that, as argued above, it is indeed related to non-normativity on the problem of sampling-based choice, even when cognitive ability is controlled. However, as I explained in Chapter 1, there is a difference between the normative strategy for a given problem (local normativity) and the normative strategy for living one's life (global normativity), which may include non-local-normative performance on some problems. Sacrificing this local-normativity might be in a person's interest if, for example, it saves time that the person might more highly value spending elsewhere – certainly, not all participants are happiest spending time and cognitive resources on

researchers' experiments! Importantly, it may also be that a strategy that provides a good fit for many problems in one's life (e.g., deliberating carefully over problems) does not work well for all problems (e.g., the current decision problem). Given the relationship of NFC to normative performance on other problems (see, e.g., Kokis et al., 2002; Macpherson & Stanovich, 2007; Stanovich & West, 1999; Toplak & Stanovich, 2002), this type of deliberation strategy is especially possible in the case of NFC. It's also possible, given high NFCs' tendency to desire to gather more information (Berzonsky & Sullivan, 1992), to accumulate consumer information (Inman et al., 1990) and to rate the usefulness of web sites according to their informational value (Kaynar & Amichai-Hamburger, 2008), that their strategy involves simply gathering as much information as possible in all scenarios. Whether or not such strategies are globally normative is up for debate and depends largely on individual preferences.

Of note, there is also one interesting finding from the studies in this thesis that relates to optimising sampling-based choice but not to the choice of a normative strategy. It is the finding from Experiment II that high NFCs are better able to maintain accurate performance across many trials; this indicates something about the ability of high NFCs (as opposed to their strategy). Such ability, however, is also important for optimising performance on sampling problems. In other words, if a person wanted to hire someone to perform these decision problems, he would be wise to consider both the ability advantage of high NFCs and their strategy disadvantage. It may also be easier to teach high NFCs to overcome their strategy deficit than it is to teach low NFCs to overcome their ability deficit. Given my finding from Experiment VI that high NFCs are able to answer knowledge-based trivia questions faster than low NFCs (with no significant difference in accuracy), it is also worthwhile to

consider that high NFCs may not just have the potential to equal low NFCs in speed, but to exceed them.

Finally, there was one finding that did not relate to the optimisation of sampling-based choice, but is interesting in its own right. This is that participant-recorded confidence growth rises according to the typical learning curve. Although discovering general principles of confidence was not part of the aims of this thesis, this finding may be useful to theorists in that field.

8.2.1. Practical implications

8.2.1.1. Teaching rationality for sampling-based choice

There are two important practical considerations that arise from my findings and the discussion above. The first is to consider whether we, as a people, would benefit from teaching students to be able to respond normatively to the problem of sampling-based choice. Certainly, given that consumer choice falls into this category, and consumer satisfaction is one of the values of Western nations, it should be seriously considered. We can also easily imagine scenarios wherein employees are required to make decisions at a certain accuracy/speed ratio; e.g., deciding which planes must land first based on level of fuel and ability to use differing runways. Recognizing that NFC is related to overspending time, but better ability to maintain performance, could be useful to employers who need to hire people who can reliably optimise these types of problem in a certain way. High NFCs may be better suited to more accuracy-focused problems, other positions (where, e.g., deductive reasoning is more important), or they may benefit from training on this problem.

8.2.1.2. Teaching meta-strategies

Baron (1985; 1993) has argued that we should teach students to develop higher NFC, calling for teaching a tendency to deliberate as a strategy that will help with many reasoning problems. The second important practical consideration is whether we might instead benefit from teaching a *meta-strategy* rather than a strategy – a meta-strategy being a strategy about how to apply one's strategies. For example, we might teach children to "stop and think" on syllogistic reasoning problems but not to deliberate so much over sampling-based decisions. If possible, this kind of teaching programme would be ideal, because, if it works, students would always apply the local-normative strategy precisely where it is needed.

8.3. Future research

In approaching the central question of this thesis – whether NFC might be related to non-normative performance for problems of sampling-based choice – I identified seven hypotheses relating to why high NFCs spend more time than low NFCs, and I pursued a broad range of studies in order to examine several of them. Naturally, some of the experimental frameworks I used could be refined for future research and, importantly, there are several other completely novel ways to test my hypotheses. Below I outline the future studies that may be of use in continuing to test these hypotheses. Naturally, it may not be necessary to employ all such studies, but I hope that future researchers will nevertheless find their enumeration useful.

8.3.1. Ability failure

The Ability Failure hypothesis, which postulates that high NFCs may simply need more time to reach accuracy levels equivalent to low NFCs, has little evidence to

support it, and there is likewise no strong theoretical reason to believe it may be true. There is thus not much point in pursuing this hypothesis further.

8.3.2. Accuracy focus

The Accuracy Focus hypothesis suggests high NFCs spend time in pursuit of greater accuracy. I focused heavily on the two most prominent features of sampling-based choice – threshold and confidence growth – in Experiments II, III and IV, in order to determine whether high NFCs have higher thresholds of evidence strength and whether their confidence takes longer to reach threshold. Both would indicate a focus on accuracy; in the former case, a higher threshold of evidence translates directly to a desire for strong evidence (and therefore accuracy), and in the latter case, a focus on accuracy should manifest in lower confidence throughout sampling, until the (presumably high) threshold is finally reached. Experiments III and IV could both be improved for future research.

In Experiment III, I presented participants with samples of fixed size; I then asked them whether they would prefer to gather more information before making a choice. I asked them to focus on accuracy when doing so, because past research had shown that people were unlikely to choose the "no choice" option (Fiedler & Kareev, 2006), which equated to my "gather more information" option, and such a study would not yield interpretable evidence if participants did not at least sometimes choose this option. However, the requested focus on accuracy may have washed out the effects of NFC. A future experiment could employ instructions requesting an equal focus on accuracy and efficiency. Given that Fiedler and Kareev did not conceptualise the "no choice" option in terms of gathering more information, it is possible that emphasis on the information gathering allowed by such an option would prevent people from avoiding it entirely. With participants less focused on accuracy, I

would expect any differences between high and low NFCs to become apparent; in particular, this hypothesis predicts that high NFCs would show the same threshold as in Experiment III, but low NFCs would lower theirs now that accuracy is not as desired.

In Experiment IV, I asked participants to track their confidence as they performed the sampling-based choice task, giving them a slider bar to do so, the end of which represented their threshold, or the moment they felt ready enough to make a choice. There are two possible useful alterations to this method. The first is to test with a broader range of evidence universes, especially easier universes, and in general to try making the task easier. Doing so might alleviate the problems discussed regarding cognitive load and therefore allow more room to see differences related to NFC. The second possible alteration is to allow the participant to report confidence numerically using a typical magnitude estimation paradigm (see, e.g., Stevens, 1975), wherein she chooses any number she likes to represent her confidence, and this can go as high as she likes. Using a numerical method would be more difficult than a slider bar, as it would require the program to stop repeatedly to ask for confidence estimates. However, it would provide an additional test of threshold to Experiment III.

One final, novel method for testing this hypothesis would be to allow participants to determine the amount of time they spend looking at each piece of evidence[18]. One could then determine if high NFCs sample the same amount of evidence as low NFCs while simply expending more time, or if they truly desire more evidence.

---

[18] I am grateful to Dr. Todd Bailey for this suggestion.

8.3.3. Extraneous thoughts

I have mentioned several hypotheses involving thinking about things other than the remembering and comparison of evidence (Improvement, Pattern-searching, Unrelated Thoughts/Boredom). One potential way to test these hypotheses is to simulate the effect of thinking about other things via imposing cognitive load. According to cognitive load theory (Sweller, 1988), imposing such a load should tax the working memory capacity of participants in much the same way as entertaining outside thoughts. Thus we should see low NFCs, under such constraints, take as much time as high NFCs not under load. Naturally, this would only test whether there are any extraneous thoughts, and not whether they are task-related. Further, this test is not ideal, as low NFCs may simply sacrifice accuracy rather than taking more time.

Second, in general it may be useful to query participants after the task with a questionnaire assessing their engagement with the task and asking whether they found themselves thinking of other things (and if so, what). Although participants often do not have accurate insights into their own mental processing (see, e.g., Nisbett & Wilson, 1977), in this case it seems a reasonable assumption that they may be able to provide some clues, as some of the strategies (e.g., pattern-searching) may be deliberate, and daydreaming is also a conscious process. Further, Jack and Roepstorff (2002) argue persuasively for the re-inclusion of introspective reports in the cognitive sciences.

8.3.3.1. Improvement

The Improvement hypothesis says that high NFCs are using their extra time in order to attempt to improve at the task. Because I found little evidence for this hypothesis, it is not the most fruitful line of inquiry. However, future tests could continue to monitor for improvement in the same way that I did.

8.3.3.2. Pattern-searching

The Pattern-searching hypothesis proposes that high NFCs are attempting to search for patterns in the data they see. One possible test for this would be to test participants on the classic probability matching problem (Yellot, 1969), in which they are presented with randomly varying stimuli of different colors. White and Koehler (2007) recently did just such a test, finding no relationship between NFC and the tendency to use a matching strategy (wherein participants try to give answers matching the probability of each possible outcome, shown by Yellot to be linked to searching for a pattern). However, they did not control for cognitive ability, which, as seen in my Experiment I, may cause the relationship to become apparent. More importantly, their probability-matching problem involved medical diagnoses rather than perceived (or explicitly stated) random samples. Thus there is certainly room for such a study involving NFC and cognitive ability.

A related hypothesis is that high NFCs are simply trying to remember more information than just the evidence seen, perhaps in an attempt to search for a broader pattern or perhaps merely because this is a global strategy they employ. This could be tested by examining the recall ability of participants at the end of the task; high NFC participants may be more likely to recall task information, such as for how many trials they chose candidate A.

8.3.3.3. Unrelated thoughts/Boredom

The Unrelated Thoughts hypothesis suggests that high NFCs are thinking about things completely unrelated to the task; i.e., they are daydreaming while devoting some set level of processing to the task. Unfortunately, it is difficult to test this hypothesis directly because there is no way to predict just what they might be thinking about (as with pattern-searching). Thus the tests I outlined at the beginning

of this section (inducing cognitive load and including a questionnaire about engagement on the task) are the best ways to test this hypothesis.

Considering that, in my discussion earlier, I rejected the Boredom hypothesis, there is little need to pursue it further. However, the proposed questionnaire could include some questions regarding boredom in order to further support its elimination.

8.3.4. Social desirability

The Social Desirability hypothesis says that high NFC participants may be spending longer sampling in order to maintain an impression of effortful thought. Although I report here some evidence supporting this hypothesis, future research could examine the context in which the participant completes the study. It's possible that impression management is less important in some contexts – e.g., informal, non-evaluative settings – and more important in other, more evaluative contexts – such as a lab in a university building, or generally places where one feels watched. Indeed, Paulhus (1984) found that participants' impression management SDR scores were higher in public as opposed to anonymous conditions, and Bateson, Nettle, and Roberts (2006) found that participants paid more into an anonymous payment box when there was an image of human eyes attached to the box. It is possible to manipulate both the evaluative nature of a context and the feeling of being watched, and both of these methods could be employed in order to compare whether high NFCs continue to sample for longer. Based on the current evidence, I would predict they would spend less time on a sampling task in a less evaluative setting than the lab.

Earlier, I suggested that social facilitation, the theory of which states that people exposed to the mere presence of others tend to exhibit speeded responses on simple tasks (e.g., Schmitt et al., 1985), may have been the reason for failing to replicate the NFC-time effect in Experiments V and VI, in which participants were

tested in a lab setting along with several others. Current theorists suggest that the reason for social facilitation is precisely evaluation apprehension (Feinberg & Aiello, 2006). It is interesting to note that this theory predicts *faster* performance in evaluative situations rather than slower. It may be that evaluation apprehension, in this case, follows a quadratic function, such that a small amount of it (e.g., from knowing one is contributing data to an experiment) results in more goal-focused behaviour (where the goal is interpreted to spend more effort/time), while larger amounts (e.g., from being surrounded by other participants in a lab) result in the usual speeded response. This is precisely what Gilliland (1980) found when he induced physiological arousal in participants via varying doses of caffeine, and arousal has long been considered a possible cause of social facilitation (Zajonc, 1965), in many cases as a by-product of evaluation apprehension. In this study, both extraverted and introverted participants exhibited quadratic trends in the speed and accuracy of their performance (although the nature of these trends differed by group). Though there is no known relationship between NFC and introversion-extraversion (Cacioppo et al., 1996), it is certainly conceivable that high NFCs too could display their own quadratic trend. Thus, the best test would employ varying levels of evaluation apprehension, in order to search for such a trend.

8.3.5. Time heuristic

I proposed the Time Heuristic hypothesis as a way of suggesting high NFCs may feel continuously spurred to spend time at this kind of task, perhaps because it is a strategy that works well on other problems. I have suggested that this feeling may be identical to the inability to disengage from the task. Thus, one clear method for testing this hypothesis is to look for a correlation between NFC and the ability to disengage from ongoing cognitive tasks.

8.3.6. New variables

In my studies, I always tested hypotheses related to experimenter-defined, as opposed to participant-defined, normativity. However, it would certainly be interesting to address whether participants are satisfied with their own performance on this type of problem; this could be done by querying satisfaction direction. Unfortunately, the task I gave to participants was not one well suited to measuring satisfaction, as a correct response would have little impact on the participant's life, and therefore differences in satisfaction would likely have been small. However, in future studies, it would certainly be possible to employ monetary rewards in order to increase this impact, or to study actual consumer choices. In this way, we could assess whether, despite time apparently wasted, high NFCs may at least report satisfaction with their choices (although we should take care to consider how much their responses are due to impression management or self-deception). Further, one important addition to the current literature may be to study both local satisfaction – that is, satisfaction relating to the choices just made – and global satisfaction – i.e., the satisfaction one has with one's strategy of choice-making after making many such choices across several years.

There are also several other individual differences variables that one could assess in relation to this task, in addition to NFC. First, it may be useful to examine the sub-scales of NFC as outlined by Tanaka, Panter, and Winborne (1988); Cognitive Persistence in particular may be more related to spending time than the other two sub-scales they suggest, Cognitive Complexity and Cognitive Confidence. Next, ctively open-minded thinking (for the current version, see Stanovich & West, 2007), which measures a person's tendency to actively seek out alternative opinions, is one thinking style thus far commonly assessed alongside reasoning tasks (e.g., Stanovich & West,

2007) and may also play a role here, although in pilot testing I did not discover that it did. Maximization (Schwartz et al., 2002), defined as a person's desire to find the best possible option, is another likely candidate, although again, I used it in pilot testing and did not find a relationship. Finally, Perfectionism is also a possibility, especially since Stoeber and Eysenck (2008) recently found that perfectionistic strivings (having high standards for oneself) was negatively correlated with accuracy/speed on a proof-reading task, and Stoeber, Chesterman, and Tarn (2010) found that that a positive correlation between perfectionistic strivings and accuracy on a letter-detection task was fully mediated by time spent, although in neither case did the researchers partial out the effects of cognitive ability. The study of several such thinking styles should not only help to conceptualise what participants are doing on the sampling-based choice task, but continue to add to the growing literature on the roles that each of these thinking styles play across various tasks.

Finally, some researchers have begun to consider the neural correlates of NFC. For example, in an EEG study, Enge, Fleischhauer, Brocke, and Strobel (2008) examined high- and low-NFC individuals who were silently counting novel auditory stimuli, finding that high NFCs showed larger P3a and P3b amplitudes in response to the novel sounds. These potentials are known to be associated with the engagement of attention; Enge et al. concluded that high NFCs were engaging in both greater involuntary and voluntary attention allocation. Others have looked at the neurological bases for employing heuristics and biases: In an fMRI study, DeMartino, Kumaran, Seymour, and Dolan (2006) reported that heuristics-followers in a typical framing task exhibited more activation in the amygdala, an area strongly associated with emotional experiences, while those who eschewed heuristics exhibited more activation in the orbitofrontal cortex and anterior cingulate, two areas of importance

in executive functions, in particular decision-making and sensitivity to reward. Given that such differences have been found, it may be illuminating to examine high and low NFCs on a sampling-based choice task; any differences noted between them on this task may be useful in determining the nature of differences in their attention and mental engagement. Further, the activation results might be compared to those from a heuristics-and-biases task at which high NFCs typically excel. A similar pattern of results across both tasks might indicate that the general approach of high NFCs is applied throughout a variety of tasks, which could mean that the very mental engagement that affords normative performance on some tasks result in non-normative performance on this task.

8.3.7. New questions

There are two main new questions that should be explored in order to better determine the practical implications of the findings discussed in this thesis. The first is whether NFC (or a meta-strategy, as I suggested earlier) can truly be taught, as proposed by Baron (1985). If not, then the entirety of the usefulness of the present line of research will be in predicting the behaviour of individuals based on their thinking styles, and we as a society would not have to decide whether to teach such strategies. However, it is certainly possible that thinking styles can be taught, and some research should be done along these lines.

The second practical question to be considered is whether the behaviour exhibited by high NFCs on this task can be changed (separately from the thinking style on the whole). As I discussed earlier, it may be important for employees to be able to perform this task effectively, and therefore the ability to train them would be useful. It would also be interesting to note whether such training could be kept separate from thinking style or if it would have an effect in changing it.

There are also several interesting questions that arise relating to the perception of time. It has already been discovered that high NFCs appear to perceive time differently when performing tasks, underestimating the amount of time spent compared to their low-NFC peers (Baugh & Mason, 1986). Might it be the case that high NFCs perceive time differently while they are performing sampling-based tasks as well? Also, how might this perception affect their performance on other tasks in which time is important, and how might it affect their judgment on tasks such as time planning, which may result in the planning fallacy? The latter case may provide a fruitful new avenue for discovery of a new non-normative behaviour (the planning fallacy) associated with NFC.

Finally, it may be beneficial to consider to what extent NFC may be considered a measure of personal *intensity*, such that the high-NFC individual generally wants *more* of everything. I have discovered here that high NFCs may want more time, and other researchers have shown they want more information (Berzonsky & Sullivan, 1992) and more emotional experiences (Maio & Esses, 2001). In accordance with this hypothesis, numerical estimation studies, if conducted, may show that high NFCs generally produce higher numbers than low NFCs. Such a finding would be of great importance because it would indicate the generalization of this tendency by high NFCs, and it could be used to predict their performance on a variety of tasks.

8.4. Conclusions

In this thesis, I have defined and discussed rationality, the role that dual process theories of cognition play in rationality, and what we can learn about both from individual differences in thinking styles, particularly the Need for Cognition. I have shown in two experiments that NFC explains unique variance in the amount of

time spent at a sampling-based choice task, even though it is not related to a significant increase in accuracy at the same task. I have found both a significant partial mediation of this NFC-time relationship by social desirability, and evidence suggesting that high NFCs prefer stronger evidence (and therefore focus more on accuracy) than low NFCs. I have argued that these findings represent non-normative behaviour on the part of high NFCs, although in other areas they display greater ability at the task, and their behaviour of course may still be globally normative. These findings of non-normativity represent the first instance of a thinking style being clearly related to both normative performance on some tasks and non-normative performance on others. They suggest caution in the interpretation that more deliberation is always better, and therefore suggest that teaching thinking should not be just about teaching a single, over-arching strategy to deliberate. Rather, we should teach a meta-strategy, one that means assessing each problem situation individually and then applying the appropriate normative strategy. We should think deeply, then, about when to think.

References

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125*(5), 591-600.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics, 55*(4), 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 929-945.

Baron, J. (1985). *Rationality and intelligence.* Cambridge: Cambridge University Press.

Baron, J. (1993). Why teach thinking? An essay. *Applied Psychology: An International Review, 42*, 191-237.

Baron, J. (2000). *Thinking and deciding.* New York: Cambridge University Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*, 412-414.

Baugh, B. T., & Mason, S. E. (1986). Need for cognition related to time perception. *Perceptual and Motor Skills, 62*, 540-542.

Bernoulli, J. (1713). *Ars conjectandi [Probability calculation].* Basel, Switzerland: Thurnisiorum.

Berzonsky, M. D., & Sullivan, C. (1992). Social-cognitive aspects of identity style: Need for cognition, experiential openness, and introspection. *Journal of Adolescent Research, 7*, 140-155.

Bills, A. G. (1934). *General experimental psychology.* New York: Longmans, Green, and Co.

Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 115-160). San Diego, CA: Academic Press.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700-765.

Bors, D. A., Vigneau, F., & Lalande, F. (2006). Measuring the need for cognition: Item polarity, dimensionality, and the relation with ability. *Personality and Individual Differences, 40*, 819-828.

Busemeyer, J., & Townsend, J. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 423-459.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological bulletin, 119*(2), 197-253.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306-307.

Cacioppo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual differences perspective. *Journal of Personality and Social Psychology, 51*, 1032-1043.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*(3), 404-431.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Cooper, J. (2007). *Cognitive dissonance: 50 years of a classic theory*. London: Sage Publications.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Curseu, P. L. (2006). Need for cognition and rationality in decision-making. *Studia Psychologica, 48*(2), 141-156.

Curşeu, P. L. (2006). Need for cognition and rationality in decision-making. *Studia Psychologica, 48*(2), 141-156.

Day, E. A., Espejo, J., Kowollik, V., Boatman, P. R., & McEntire, L. E. (2007). Modeling the links between need for cognition and the acquisition of a complex skill. *Personality and Individual Differences, 42*, 201-212.

DeMartino, B., Kumaran, D., Seymour, B., & Dolan, R. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313*, 684-687.

Dijksterhuis, A., Bos, M., Nordgren, L., & van Baaren, R. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*, 1005-1007.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33*(3), 261-268.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*, 171-185.

Einhorn, H., & Hogarth, R. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32*, 53-88.

Enge, S., Fleischhauer, M., Brocke, B., & Strobel, A. (2008). Neurophysiological measures of involuntary and voluntary attention allocation and dispositional differences in Need for Cognition. *Personality and Social Psychology Bulletin, 34*(6), 862-874.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin, 30*(4), 447-460.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*(4), 382-389.

Evans, J. S. B. T., & Over, D. (2004). *If.* Oxford: Oxford University Press.

Evans, L., & Buehner, M. J. (in press). Small samples do not cause greater accuracy - But clear data may cause small samples. Comment on Fiedler and Kareev (2006). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Fantino, E., Kanevsky, I. G., & Charlton, S. R. (2005). Teaching pigeons to commit base-rate neglect. *Psychological Science, 16*(10), 820-825.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness - The development and correlates of a new scale. *Journal of Personality Assessment, 50*(1), 4-17.

Feinberg, J. M., & Aiello, J. R. (2006). Social facilitation: A test of competing theories. *Journal of Applied Social Psychology, 36*(5), 1087-1109.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58*, 203-210.

Fiedler, K., & Kareev, Y. (2006). Does decision quality (always) increase with the size of information samples? Some vicissitudes in applying the law of large numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 883-903.

Fitts, P. M., & Posner, M. I. (1967). *Human performance.* Ann Arbor, MI: Brooks/Cole Pub. Co.

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin, 36*(1), 82-96.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15*(2), 105-128.

Garrett, H. E. (1922). A study of the relation of accuracy to speed. *Archives of Psychology, 56*, 1-104.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107-143.

Gilliland, K. (1980). The interactive effect of introversion-extraversion with caffeine induced arousal on verbal performance. *Journal of Research in Personality, 14*, 482-492.

Goodie, A. S., & Fantino, E. (1995). An experientially derived base-rate error in humans. *Psychological Science, 6*, 101-106.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411-435.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116*(2), 454-461.

Hartl, J. A., & Fantino, E. (1996). Choice as a function of reinforcement ratios in delayed matching to sample. *Journal of Experimental Analysis of Behavior, 66*, 11-27.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review, 18*, 186-201.

Herbranson, W. T., & Schroeder, J. (2010). Are birds smarter than mathematicians? Pigeons (Columba livia) perform optimally on a version of the Monty Hall Dilemma. *Journal of Comparative Psychology, 124*(1), 1-13.

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind* (pp. 209-235). New York: Oxford University Press.

Hick, W. F. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Hinson, J. M., & Staddon, J. E. R. (1983). Matching, maximizing, and hill-climbing. *Journal of the Experimental Analysis of Behavior, 40*, 321-331.

Inman, J. J., McAlister, L., & Hoyer, W. D. (1990). Promotion signal: Proxy for a price cut? *Journal of Consumer Research, 17*, 74-81.

Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. (1956). Tests of two theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology, 51*(4), 261-268.

Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: From stimulus response to script-report. *Trends in Cognitive Sciences, 6*(8), 333-339.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6839-6833.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217.

Kaynar, O., & Amichai-Hamburger, Y. (2008). The effects of Need for Cognition on Internet use revisited. *Computers in Human Behavior, 24*(2), 361-371.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 1, pp. 192-238). Lincoln: University of Nebraska Press.

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology, 89*(3), 470-485.

Klaczynski, P. A., & Lavallee, K. L. (2005). Domain-specific identity, epistemic regulation, and intellectual ability as predictors of belief-biased reasoning: A dual-process perspective. *Journal of Experimental Child Psychology, 92*, 1-24.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning biases. *Psychology and Aging, 15*(3), 400-416.

Klayman, J., Soll, J. B., Juslin, P., & Winman, A. (2006). Subjective confidence and the sampling of knowledge. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition.* Cambridge: Cambridge University Press.

Kokis, J., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*, 26-52.

Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences, 17*, 115-127.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India 2*(1), 49-55.

Maio, G. R., & Esses, V. M. (2001). The Need for Affect: Individual differences in the motivation to approach or avoid emotions. *Journal of Personality, 69*(4), 583-615.

Martin, B. A. S., Sherrard, M. J., & Wentzel, D. (2005). The role of sensation seeking and need for cognition on web-site evaluations: A resource-matching perspective. *Psychology & Marketing, 22*, 109-124.

Martinez, M. E. (2000). *Education as the cultivation of intelligence.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Merkle, E. C., Sieck, W. R., & Van Zandt, T. (2008). Response error and processing biases in confidence judgment. *Journal of Behavioral Decision Making, 21*(4), 428-448.

Merkle, E. C., & Van Zandt, T. (2006). An application of the Poisson race model to confidence calibration. *Journal of Experimental Psychology: General, 135*(3), 391-408.

Newton, E. J., & Roberts, M. J. (2005). The window of opportunity: A model for strategy discovery. In M. J. Roberts & E. J. Newton (Eds.), *Methods of thought: Individual differences in reasoning strategies.* Hove: Psychology Press.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625-631.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 251-259.

Oaksford, M., & Chater, N. (1998a). A revised rational analysis of the selection task: Exceptions and sequential sampling. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 372-398). Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (1998b). An introduction to rational models of cognition. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 1-18). Oxford: Oxford University Press.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.

Perkins, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence.* New York: Free Press.

Pleskac, T. J., & Busemeyer, J. R. (2009). Two-stage dynamic siganal detection theory: A dynamic and stochastic theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864-901.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116*(1), 59-83.

Raven, J. C. (1962). *Advanced Progressive Matrices (Set II)*. London: H.K. Lewis & Co.

Raven, J. C., Court, J. H., & Raven, J. (1977). *Manual for Advanced Progressive Matrices (Sets I & II)*. London: H.K. Lewis & Co.

Roberts, M. J., & Newton, E. J. (2002). Inspection times, the change task, and the rapid-response selection task. *Cognition, 31*, 117-140.

Robinson, R. B., Cox, C. D., & Odom, K. (2005). Identifying outliers in correlated water quality data. *Journal of Environmental Engineering, 131*(4), 651-657.

Schmitt, B. H., Gilovich, T., Goore, N., & Joseph, L. (1985). Mere presence and social facilitation: One more time. *Journal of Experimental Social Psychology, 22*(3), 242-248.

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica, 27*, 143-153.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology, 83*(5), 1178-1197.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*(1), 33-51.

See, Y. H. M., Petty, R. E., & Evans, L. M. (2009). The impact of perceived message complexity and need for cognition on information processing and attitudes. *Journal of Research in Personality, 43*(5), 880-889.

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science, 19*, 368-373.

Smith, P. L., & Vickers, D. (2009). Modeling evidence accumulation with partial loss in expanded judgment. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 797-815.

Sojka, J. Z., & Giese, J. L. (2001). The influence of personality traits on the processing of visual and verbal information. *Marketing Letters, 12*, 1.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Stanovich, K., & West, R. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188.

Stanovich, K. E. (2008). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55-88). Oxford: Oxford University Press.

Stanovich, K. E. (2009). Rational and irrational thought: The thinking that IQ tests miss. *Scientific American Mind, 20*(6), 34-39.

Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching and human development* (pp. 195-237). New York: Springer.

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. *Advances in Child Development and Behavior, 35*, 251-285.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188.

Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative/descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology, 38*, 349-385.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning, 13*, 225-247.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2010). Individual differences as essential components of heuristics and biases research. In K. Manktelow, D. Over & S. Elqayam (Eds.), *The science of reason: A festschrift for Jonathan St. B. T. Evans* (pp. 335-396): Psychology Press.

Stevens, S. S. (1975). *Psychophysics*. USA: John Wiley & Sons, Inc.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment, 78*, 370-389.

Stoeber, J., Chesterman, D., & Tarn, T. (2010). Perfectionism and task performance: Time on task mediates the perfectionistic strivings–performance relationship. *Personality and Individual Differences, 48*(4), 458-462.

Stoeber, J., & Eysenck, M. W. (2008). Perfectionism and efficiency: Accuracy, response bias, and invested time in proof-reading performance. *Journal of Research in Personality, 42*(6), 1673-1678.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Tanaka, J. S., Panter, A. T., & Winborne, W. C. (1988). Dimensions of the need for cognition: Subscales and gender differences. *Multivariate Behavioral Research, 23*, 35-50.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond.* Oxford: Oxford University Press.

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 94*(1), 197-209.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293-315.

Verplanken, B., Hazenberg, P., & Palenéwen, G. R. (1992). Need for cognition and external information search effort. *Journal of Research in Personality, 26*(2), 128-136.

Vickers, D., Burt, J., Smith, P., & Brown, M. (1985). Experimental paradigms emphasising state or process limitations: I. Effects on speed-accuracy trade-offs. *Acta Psychologica, 59*, 129-161.

Von Neumann, J., & Morganstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology 1*. Harmondsworth, England: Penguin Books Ltd.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Wedell, D. H. (in press). Probabilistic reasoning in prediction and diagnosis: Effects of problem type, response mode, and individual differences. *Journal of Behavioral Decision Making*, 1-23.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930-941.

White, C. M., & Koehler, D. J. (2007). Choice strategies in multiple-cue probability learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 757-768.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67-85.

Woodworth, R. S. (1899). Accuracy of voluntary movement. *Psychological Review Monographs*, 27-54.

Yellot, J. I., Jr. (1969). Probability learning with noncontingent success. *Journal of Mathematical Psychology, 6*, 541-575.

Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269-274.

Appendixes

Appendix A: Need for Cognition scale

Below is the Need for Cognition scale, short form (Cacioppo et al., 1984). Items marked with an asterisk are reverse-scored.

1. I would prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun.*
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.*
5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.*
6. I find satisfaction in deliberating hard and for long hours.
7. I only think as hard as I have to.*
8. I prefer to think about small, daily projects to long-term ones.*
9. I like tasks that require little thought once I've learned them.*
10. The idea of relying on thought to make my way to the top appeals to me.
11. I really enjoy a task that involves coming up with new solutions to problems.
12. Learning new ways to think doesn't excite me very much.*
13. I prefer my life to be filled with puzzles that I must solve.
14. The notion of thinking abstractly is appealing to me.
15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.*
17. It's enough for me that something gets the job done; I don't care how or why it works.*
18. I usually end up deliberating about issues even when they do not affect me personally.

Appendix B: Instructions for Experiments I and II

Below are the instructions for the choice task used in Experiments I and II, which are by and large a direct translation of Fiedler and Kareev's (2006) instructions, with some minor clarifications and additional instructions regarding the particularities of the computer program used. Text has been bolded here to indicate accuracy and efficiency instructions, the order of which changed by participant (see Experiment I, Method); the below represents one of the two possible orders.

Instructions

Imagine you are a decision maker in a personnel department. A new department has been set up in your company, and numerous posts need to be filled.

In order to facilitate this process, job candidates already took part in an assessment centre. This means that they had to solve various tasks, either in a group, or as individuals, to demonstrate their ability for the positions in question. While they participated, they were filmed.

These films in turn have since been shown to a large number of independent jurors, who then made a judgment about the suitability of a candidate (positive or negative). All these judgments were recorded and saved in a database. You can now sample judgments about two candidates at a time from this database, in order to decide which of them is better - as measured by the total number of juror judgments.

However, due to time constraints, it is not possible to access all data relative to both applicants. You thus have to rely on a sample of judgments. In doing so, you determine the size of the sample (i.e. the number of juror judgments) yourself. You will be allowed to discard any trial without making a decision if you feel you cannot decide between the two candidates.

Both **accuracy** and **efficiency** are important in working on this task. This

means that within the time available, you should **compare as many pairs of candidates as possible**, but at the same time **try not to make any mistakes (i.e. always determine the better candidate as defined by the judgments of the jurors)**.

Each trial works like this: The judgments about both respective candidates appear together on the screen. The left side of the screen (shaded in turquoise) always represents candidate A, the right side (shaded in orange), candidate B. While you keep the button at the bottom of the screen depressed, judgments about both candidates, drawn randomly from the sample, will appear on the screen. A positive impression about a candidate will be represented as a smiley face, a negative impression as a frowny face. The longer you press the button, the more juror judgments will appear on the screen, each time assigned to one of the two candidates. Let go of the button only once you think that you have seen enough judgments to decide on one of the two candidates. You will then see a small menu which will give you three options:

Choose Candidate A

Choose Candidate B

No Choice - Don't choose either of the candidates and discard the trial

Remember, both **efficiency** and **accuracy** are required. You should **preferably only make correct decisions**, but **try to make as many decisions as possible**.

At the end of the experiment you will be given feedback about how you scored.

You will begin with a practice block of 2 trials to give you a feel for the task before you begin the experiment. After each trial, there will be a short pause before the next trial begins. Don't worry - it isn't as complicated as it sounds. Give it a try!

NOTE: Be sure to click and HOLD the button - if you just click, you will lose your chance to see more judgments for that trial.

Appendix C: Instructions for Experiment III

Below are the instructions for the choice task used in Experiment III. Note that, since this experiment was delivered immediately after the choice task in Experiment II, participants were already familiar with the task, having received the task instructions for Experiment II (see Appendix B).

Instructions

After a while, your boss enters and tells you that he wants you to change your focus: he wants you to be as ACCURATE as possible. Thus only make a decision when you are REALLY CERTAIN you are correct. And remember: the judges are people and they have been known to make mistakes.

Also, you decide now to take a different approach to gathering data on the candidates, and you arrange to look at the judgments in small groups rather than one at a time. You can make a choice right away for a candidate, or you can decide to hold off and gather more data on them later. Remember, you are to choose the *better* candidate, and you are to be as ACCURATE as possible.

Appendix D: Instructions for Experiment IV

Below are the instructions for Experiment IV. The first half of the instructions was identical to that in Experiments I and II (see Appendix B) and has been omitted. Text has been bolded to indicate accuracy and efficiency instructions, the order of which changed by participant (see Experiment I, Method); the below represents one of the two possible orders.

Instructions

Each trial works like this: The judgments about both respective candidates appear together on the screen. The left side of the screen (shaded in turquoise) always represents candidate A, the right side (shaded in orange), candidate B. Once you have started to move the slider on screen, judgments about both candidates, drawn randomly from the sample, will appear on the screen. A positive impression about a candidate will be represented as a smiley face, a negative impression as a frowny face.

Importantly, we are interested in tracking your CONFIDENCE while you do this task – that is, how likely do you think it is you will choose the better candidate? Thus we ask you to CONTINUE TO MOVE THE SLIDER in line with your confidence that you will choose correctly. Consider the end of the slider scale to be the maximum confidence needed to make a choice – when you reach the end, judgments will stop and you will be asked to make your decision. If you find you are never very confident but want to stop anyway, you may press the Stop Now button.

If the slider reaches its end or you press the Stop Now button, you will then see a small menu which will give you three options:

Choose Candidate A

Choose Candidate B

No Choice - Don't choose either of the candidates and discard the trial

Remember, both **efficiency** and **accuracy** are required. You should **preferably only make correct decisions**, but **try to make as many decisions as possible**.

You will begin with a practice block of 3 trials to give you a feel for the task before you begin the experiment. After each trial, there will be a short pause before the next trial begins. And remember, you only have to choose the *better* of the two candidates.

Don't worry - it isn't as complicated as it sounds. Give it a try!

Appendix E: Instructions for Experiment V

Below are the focusing instructions for Experiment V's Accuracy and Speed conditions; they were placed in the same two locations as the accuracy/efficiency instructions in Experiments I and II (see Appendix B). The remainder of the instructions are identical to those in Experiments I and II and have been omitted. All instructions for the Control condition were identical to those in Experiments I and II and have thus also been omitted.

Instructions for Accuracy Condition

Please focus on accuracy in working on this task. This means that you should try not to make any mistakes (i.e. always determine the better candidate).

Remember, please focus on accuracy. You should preferably only make correct decisions.

Instructions for Speed Condition

Please focus on efficiency in working on this task. This means that you should try to get through all the candidate pairs as quickly as possible.

Remember, please focus on efficiency. You should always try to respond as quickly as possible.

Appendix F: Instructions for Experiment VI

Below are the instructions for Experiment VI. Block 1's instructions (for Sampling Questions) are given first, followed by Block 2's instructions (for Knowledge Questions). In each case, I have listed all three possible endings to the instructions, one per group (Control, Accuracy, or Speed). In the case of the Control instructions, accuracy and efficiency related instructions have been bolded. The order of such words was changed per participant in the same manner as in Experiment I (see Experiment I, Method); the below represents one of the two possible orders.

Instructions – Sampling Questions

For this task, you will see a series of random numbers between 1 and 100, presented one after another. The final number will remain on screen. You will then be asked a question whose answer will always be a percentage. For example, "What percentage of people have blood type O?"

You must judge, first, whether you think the answer is above or below the number on screen. Answer as soon as you think you know, and please only choose once. (E.g., do not switch from choosing Above to choosing Below.)

You will then be asked what you think the actual answer is. Fill this in in the box provided.

It is very important, however, that you try to answer the first question ("Above or below?") as soon as you know it – then go on to try to judge the exact amount of the answer.

CONTROL

Finally, both **accuracy** and **efficiency** are important in this task. That is, you should **move through the trials as quickly as possible** and **preferably only make correct decisions.**

ACCURACY

Finally, please focus on accuracy in this task. That is, you should always try to get the answer right.

SPEED

Finally, please focus on efficiency in this task. That is, you should always try to respond as quickly as possible.

Instructions – Knowledge Questions

This task is exactly the same as the last task, except the questions you will be asked are numerical rather than percentage based. For example, "How many cards are in a pack?"

Again, the numbers you see will be random and between 1 and 100, the final number remaining on screen. The correct answer is also always between 1 and 100.

Please still answer the first question ("Above or below?") as soon as you think you know it, without changing your choice once you've made it. Then go on to try to judge the exact amount of the answer.

CONTROL

Both **accuracy** and **efficiency** are still important in this task. That is, you should **move through the trials as quickly as possible** and **preferably only make correct decisions.**

ACCURACY

Please still focus on accuracy in this task. That is, you should always try to get the answer right.

SPEED

Please still focus on efficiency in this task. That is, you should always try to respond as quickly as possible.