2015

Believe It or Not: Examining the Case for Intuitive Logic and Effortful Beliefs

Stephanie Howarth

PHD

Believe It or Not: Examining the Case for Intuitive Logic and Effortful Beliefs

Howarth, Stephanie

Award date:
2015

Awarding institution:
University of Plymouth

Link to publication in PEARL

Believe It or Not: Examining the Case for Intuitive Logic and Effortful Beliefs

By Stephanie Howarth

A thesis submitted to Plymouth University in partial fulfilment for the degree of

Doctor of Philosophy

December 2014

School of Psychology

Faculty of Health and Human Science

Plymouth University

Abstract

Believe it or not: Examining the case for intuitive logic and effortful beliefs

By Stephanie Howarth

The overall objective of this thesis was to test the *Default Interventionist* (DI) account of belief-bias in human reasoning using the novel methodology introduced by Handley, Newstead & Trippas (2011). DI accounts focus on how our prior beliefs are the intuitive output that bias our reasoning process (Evans, 2006), whilst judgments based on logical validity require effortful processing. However, recent research has suggested that reasoning on the basis of beliefs may not be as fast and automatic as previous accounts claim. In order to investigate whether belief based judgments are resource demanding we instructed participants to reason on the basis of both the validity and believability of a conclusion whilst simultaneously engaging in a secondary task (Experiment 1 - 5). We used both a within and between subjects design (Experiment 5) examining both simple and complex arguments (Experiment 4 – 9). We also analysed the effect of incorporating additional instructional conditions (Experiment 7 – 9) and tested the relationships between various individual differences (ID) measures under belief and logic instruction (Experiment 4, 5, 7, 8, & 9).

In line with Handley et al.'s findings we found that belief based judgments were more prone to error and that the logical structure of a problem interfered with judging the believability of its conclusion, contrary to the DI account of reasoning. However, logical outputs sometimes took longer to complete and were more affected by random number generation (RNG) (Experiment 5). To reconcile these findings we examined the role of Working Memory (WM) and Inhibition in Experiments 7 – 9 and found, contrary to Experiment 5, belief judgments

were more demanding of executive resources and correlated with ID measures of WM and inhibition.

Given that belief based judgments resulted in more errors and were more impacted on by the validity of an argument the behavioural data does not fit with the DI account of reasoning. Consequently, we propose that there are two routes to a logical solution and present an alternative Parallel Competitive model to explain the data. We conjecture that when instructed to reason on the basis of belief an automatic logical output completes and provides the reasoner with an intuitive logical cue which requires inhibiting in order for the belief based response to be generated. This creates a Type1/Type 2 conflict, explaining the impact of logic on belief based judgments. When explicitly instructed to reason logically, it takes deliberate Type 2 processing to arrive at the logical solution. The engagement in Type 2 processing in order to produce an explicit logical output is impacted on by demanding secondary tasks (RNG) and any task that interferes with the integration of premise information (Experiments 8 and 9) leading to increased latencies. However the relatively simple nature of the problems means that accuracy is less affected.

We conclude that the type of instructions provided along with the complexity of the problem and the inhibitory demands of the task all play key roles in determining the difficulty and time course of logical and belief based responses.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank Professor Simon Handley and Dr Clare Walsh for all the support and guidance they have provided me with over the last three years. I could not have asked for better supervisors.

I would like to thank my parents, Ron & Fili and my brothers Mark & Chris for all their love and reassurance and for putting up with me along the way.

Thank you to Sam Johnson for all the meals, the drinks, the music, the love and support.

And last but certainly not least thank you to Michelle Neale for the encouragement and words of wisdom from the very beginning and to all of you who have kept my spirits high and entertained me on this journey. I am forever grateful.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

Work submitted for this research degree at Plymouth University has not formed part of any other degree either at Plymouth University or at another institute.

This research degree was financed with the aid of a studentship from the Social Science Doctoral Training Centre.

This thesis is the result of the author's own investigations, carried out under the guidance of her supervisory team. Other sources are explicitly referenced.

Publications:

Howarth, S., Handley, S.J. & Walsh, C. (under review). *The Logic-Bias Effect: The role of effortful processing in the resolution of belief-logic conflict.*

Howarth, S., Handley, S.J. & Walsh, C. (in preparation). *The Role of Executive Functions in Conditional Reasoning.*

Conference Presentations:

Howarth, S. (2012). Secondary Loads and Conditional reasoning. *Talk presented at the Thinking and Reasoning seminar, Plymouth University, Plymouth.*

Howarth, S. (2012). The Effect of Distinct Secondary Loads on Conditional reasoning. *Talk presented at the 4th School of Psychology Postgraduate & Staff Conference: "Beyond the Science of the Mind". Plymouth University, Plymouth.*

Howarth, S. (2012). Are Belief Biased by Logic? The Effect of Distinct Secondary Loads on Conditional reasoning. *Talk presented at (ITC) International Conference on Thinking & Reasoning, Birkbeck University, London.*

Howarth, S. (2012). Are Belief Biased by Logic? The Effect of Distinct Secondary Loads on Conditional reasoning. *Talk presented at PsyPAG, Northumbria University, Newcastle.*

Howarth, S. (2013). The effect of Random Number Generation and Complexity on Conditional Reasoning. *Poster Presented at CogSci Conference, Humboldt-Universtat, Berlin, Germany.*

Howarth, S. (2013). The effect of Random Number Generation and Complexity on Conditionals. *Poster Presented at BPS Cognitive and Developmental section conference. University of Reading, Reading.*

Word count for the main body of the thesis: 47,587

Signed

15th December 2014

Chapter 1: Literature Review

1.1 Introduction to the Review

Throughout the reasoning literature, many researchers have examined response accuracy scores to problems in which the believability of the conclusion conflicts with the logical status of the argument (see Sá, West & Stanovich, 1999), take the following example:

All plants need water

Roses need water

Therefore, roses are plants.

In the case presented above the invalid conclusion is often endorsed because it represents what is believed to be true (Evans, Barston & Pollard, 1983). However if the content of the argument is changed to the following:

All plants need water

Humans need water

Therefore, Humans are plants.

Belief in the conclusion is no longer in conflict with its logical status, making the conclusion easier to reject (Evans et al., 1983). These examples help illustrate the impact that beliefs in a conclusion can have over the ability to interpret the validity of an argument. However, the influence of our beliefs is not confined to reasoning, consider the following problem:

Jack is 34 years old and from Kent. He studied Marine Biology at Plymouth University and wrote his thesis on the conservation of vulnerable marine life. As a student he worked part time at the city Aquarium.

Which statement is more probable; that a) Jack enjoys drinking Malt Whisky or that b) Jack is a lecturer in Marine Biology and enjoys drinking Malt Whisky? According to Tversky and Kahneman (1983) those who give the answer b, are using the information representative of the person from the description or the `representative heuristic' as opposed to giving the statistically more likely outcome. The Conjunction Fallacy, as this example is referred to, highlights the conflict we experience between a response based on what we believe is true (or more representative of the information presented) to one that requires more effortful processing. In this case to establish that a conjunction of two characteristics cannot be more probable than one characteristic on its own (P (A + B) $\leq$ PA - conjunction rule). Nevertheless, even though I know that `a' is more probable, there remains a strong intuition that Jack is a whisky drinking Marine Biologist.

The Dual Process Theory (DPT) provides a theoretical framework to explain the conflict experienced between two systems, with distinct cognitive processes, generating conflicting responses. To test their predictions dual process theorists often use some of the tasks described above and typically instruct participants to reason logically. In this thesis I will examine the accuracy and latency of judgments when participants are instructed to evaluate either the logical validity or believability of presented conclusions. The objective of this thesis is to test traditional dual process accounts of belief-bias by examining the impact of secondary tasks and executive processes on the performance of judgments made under belief or logic instruction.

This chapter begins with an overview of Dual Process Theories (DPT) followed by an introduction to the Default Interventions (DI) DPT of reasoning. We review some paradigms used to test the DI account, namely the belief-bias paradigm and examine some of the research interpreted as providing empirical support for the DI DPT, covering literature on individual differences and neuropsychology. This leads on to some general critique of DPT followed by a more specific critique of the DI account from research on conflict detection. We will also consider the implications of research that proposes an intuitive route for detecting logical validity. The review finishes with discussion of an alternative Parallel Competitive (PC) dual process account and the role of inhibition in reasoning. Finally, the chapter concludes with a rationale for the experimental work presented in this thesis.

1.2 Introduction to Dual Process Theories

Classic DPT of deductive reasoning suggests that there are two systems producing responses to reasoning problems. The first, often referred to as *System 1,* is described as fast, heuristic and dependant on context for the production of intuitive conclusions. The more controlled, analytic system, often referred to as *System 2,* is context independent and can deliver logical inferences using explicit information only (Stanovich & West, 2000).

This model has been extended to several areas of research including learning (e.g. Reber, 1996), attention (Shiffrin & Schneider, 1977), reasoning (Evans, 2003), decision making (Kahneman & Frederick, 2002) and social cognition (Chaiken & Trope, 1999). These distinct cognitive systems have been defined in different ways throughout the literature; for example, Epstein (1994) described one system as experiential and the other as rational. Sloman (1996) defined the systems as associative versus rule-based.

Stanovich (1999) labelled them System 1 (heuristic and associative) and System 2 (rule-based). However, the processing features within each system have been defined as sharing some fundamental characteristics. Evans (2009) suggested that these processes can be classified as Type 1: fast, automatic, unconscious and effortless by nature and Type 2: slow, controlled, conscious and effortful. Type 1 processes are described as relatively undemanding of computational capacity and based on personal experience and associations (Verschueren, Schaeken & d'Ydewalle, 2005). Type 2 processes, on the other hand, are necessary for hypothetical thinking (Evans & Over, 1996) and are demanding of working memory.

In the Conjunction fallacy problem presented earlier, using a representativeness heuristic to arrive at the answer would be a Type 1 process, whereas applying the conjunction rule would require some Type 2 processing. Type 1 and Type 2 processes are assumed to play significant roles in reasoning, judgment and decision making and the conflict between these two processes can explain why for example, we can simultaneously demonstrate stereotypes in measures of implicit attitudes (such as the IAT), but show no evidence when asked about our explicit attitudes to different groups. A stereotyped response would be considered an automatic, Type 1 output that requires inhibiting or overriding in favour of a more considered logically deduced Type 2 output. This is known as the *Default Interventionist* (DI) dual process account (Evans; 2003; 2007), which has been developed in part to explain other experimental phenomena such as the belief-bias effect which will be discussed further in section 1.2.2.

1.2.1 Default Interventionist Dual Process Accounts of Reasoning

The DI DPT posited by Evans (2008) defines Type 1 processes as generating default, automatically cued responses (Evans, 2006). These responses are based upon beliefs, learned associations or stereotypes and can be resisted or intervened on by more conscious, deliberative Type 2 processing, if a conflict between the two responses is detected. In order to produce an alternative response, cognitive effort and working memory resources are required to inhibit the Type 1 output (Handley, Capon, Beveridge, Dennis & Evans, 2004). However, overriding a Type 1 response would also depend on adequate cognitive ability, the application of relevant instructions and having sufficient time available for more reflective thinking (Evans & Curtis-Holmes, 2005). Consequently, the initial intuitive response to a problem is often accepted (Stanovich, 2009a).

There is a considerable amount of experimental research that has been construed as evidence for the DI dual processing account which we will address in more detail in the next section. First we will consider the belief-bias effect as one of the key methods offering empirical support for dual process accounts in reasoning (Evans, Barston, & Pollard, 1983; Stanovich, 1999; Wilkins, 1929).

1.2.2 Belief Bias

Traditionally, belief-bias is the label given to the phenomena of assessing the conclusion of a deductive inference on the basis of personal belief, rather than the logical validity of the argument. Take the following example;

If a child is crying, then it is happy,

Suppose a child is crying,

Does it follow that the child is happy?

According to propositional logic a valid conclusion would be that the child is happy, however people's beliefs often influence the inferences they draw. In the example above people will frequently judge that the conclusion does not follow because it is inconsistent with the belief that a child is sad when it cries (Evans, Handley, Neilens &

Over, 2010). Social psychology provides much evidence on how prior opinion biases the evaluation of communication (Biek, Wood & Chaiken, 1996; Dole & Sinatra, 1998) whilst the ability to resist these beliefs is seen as a fundamental part of critical thinking (Sa, West & Stanovich, 1999).

Evans, Barston & Pollard's (1983) research on belief-bias with syllogisms found three common effects from the manipulation of belief and logic. Evans et al. presented participants with four types of syllogisms; valid-believable, valid-unbelievable, invalid-believable and invalid-unbelievable (see Table 1.1) and asked participants to judge conclusion validity (i.e. whether the conclusion necessarily followed from the premises). The three common effects were; a main effect of logic, showing competence in deductive reasoning, a main effect of belief, indicating the strong influence beliefs have over judgments and an interaction between belief and logic, demonstrating how we rely on our beliefs more when the conclusion of an arguments is invalid.

Table 1.1
Overall percentages for the acceptance of the conclusion as a function of logical validity and believability, taken from Evans et al., (1983).

	Believable	*Unbelievable*
Valid	No police dogs are vicious. Some highly trained police dogs are vicious. Therefore, some highly trained dogs are not police dogs. 89%	No nutritional things are inexpensive. Some vitamins are inexpensive. Therefore, some vitamins are not nutritional. 56%
Invalid	No addictive things are inexpensive. Some cigarettes are inexpensive. Therefore, some addictive things are not cigarettes. 71%	No millionaires are hard workers. Some rich people are hard workers. Therefore, some millionaires are not rich people. 10%

Research on belief-bias has provided significant insight into the way in which people reason when belief and logic conflict (Evans et al., 1983; Wilkins, 1929). Particularly, research has shown that whilst heuristics such as a belief heuristic or representativeness can be economically effective when making judgments, they often lead to predictable errors in both the laymen and the expert population (Tversky and Kahneman, 1974).

There is research to suggest that some individuals can follow the rules of logic when high in intellectual ability (Newstead, Handley, Harley, Wright & Farrelly, 2004), when emotionally charged and when given specific instructions. Goel and Vartanian (2011) showed that administering syllogisms with politically incorrect negative content (i.e. the justification of rape) actually weakened the belief-bias effect. Evans, Newstead, Allen & Pollard (1994) demonstrated that providing participants with detailed instructions with specific reference to the logical necessity of the argument also helped to reduce belief-bias. However the effect was not completely eliminated, corroborating the aforementioned research which illustrates our propensity to error in favour of prior knowledge.

Neuropsychological studies on belief-bias offer further support for a dual mechanism theory and evidence for distinct brain regions involved in belief and logic based reasoning. More specifically, Goel & Dolan (2003) posit that belief-bias may be influenced by brain areas associated with emotional processing.

The dual process theory was developed in part to explain the belief-bias effect, equally the effect is used to support dual processing; more specifically DI dual processing which emphasises the Type 1 nature of responses based on prior knowledge that conflict with the Type 2 processing required to process the underlying structure of the problem. Furthermore, according to the DI account, Type 1 processes are quick and effortless and precede the slower Type 2 processes which are reliant on working memory resources (Evans, 2003; Handley, Newstead & Trippas, 2011). In the next section we will review additional research interpreted as empirical support for the DI dual processing account which included research employing speeded tasks, individual differences and neuropsychological effects of the belief-bias phenomenon.

1.3 Empirical Support for DI Dual Processing

As mentioned in the introduction to DPT, dual systems or dual processing has extended across many areas of research and in general the theories associate similar attributes to Type 1 and Type 2 processes. For example, the evolutionary old nature of Type 1 processing which is rapid, automatic and requires little effort compared to, analytical

Type 2 processes, which are conscious, more controlled and are required for logical reasoning (Evans, 2008).

Table 1.2
Examples of support for Dual Processing

Secondary Task	**Developmental**
De Neys (2006) instructed participants to complete syllogistic reasoning problems whilst loading working memory with a secondary task. Results showed that WM load only significantly impacted on logic based reasoning.	Kokis, Macpherson, Toplak, West, and Stanovich (2002) examined children between the ages of 10 and 13years and showed that analytical, Type 2 processing correlated with age and cognitive ability.
Individual Differences	**Neuropsychology**
Type 2 functioning is known to require working memory, which in turn, is highly correlated reasoning ability (Capon et al. 2003). Consistent with this, Stanovich and West (2000) demonstrated that normatively correct responses correlated with higher cognitive ability. Additionally, Newstead et al. (2004) further demonstrated that intellectual ability was a good predictor of logical performance in syllogistic reasoning.	Goel et al. (2000) examined fMRI data on reasoning tasks to show that distinct parts of the brain are responsible for two different kinds of reasoning. Goel & Dolan (2003) demonstrated that the left temporal lobe system was activated during belief based reasoning, whilst the right lateral prefrontal cortex was activated when inhibition, associated with belief-bias, was required in order to complete a logical task.

Table 1.2 illustrates some of the areas of research offering empirical support for dual processing. As stated previously, the DI account emphasises the quick and effortless nature of belief based reasoning and the idea that responses based on prior knowledge would come before a logical output has time to complete. Evans and Curtis-Holmes (2005) examined this idea with a syllogistic reasoning study carried out under limited time and showed that increased time pressure increased belief based reasoning. This finding is consistent with the view that belief based responses are available early and are consequently more common when time limits are imposed.

According to Evans (2009) Type 2 processes are slower, effortful and pull on working memory (WM) resources, therefore the Type 2 processes required for reasoning about the structure of an argument should be affected by boosting the demands on WM. Research investigating the effect of increasing cognitive load and thus limiting WM resources supported this notion by magnifying the belief-bias effect (De Neys, 2006; Quayle & Ball, 2000).

For example, De Neys' (2006) research reinforced the claim that the heuristic system (Type 1 processes) operates automatically, whilst the analytic system (Type 2 processes) pulls on executive resources. De Neys presented participants with a dot memory task, consisting of a dot pattern which the participants were required to remember prior to completing a syllogistic reasoning task, and then reproduce in an empty matrix. The findings offered support to the DI dual process framework by demonstrating that burdening executive resources increased the rate of belief-bias. This was interpreted as; increasing the load on WM only affects the analytic system responsible for overriding the heuristic response to belief-logic conflict questions.

Similarly, Quayle and Ball (2000) established an association between spatial recall abilities and the belief-bias effect, showing that those with a lower spatial span produced more belief-bias, offering additional support to the DI account. However, De Neys (2006) also argued that it is not that individuals differ in their motivation to reason logically, more that those with higher WM span are more successful at completing the analytic process required to reason logically, or as Quayle and Ball (2000) explain; at times the demands of judging logical necessity is simply too great if there are insufficient resources available.

Thompson, Prowse Turner & Pennycook (2011) argue that a monitoring mechanism governed by a `feeling of rightness' for an answer is an important addition to a complete dual process theory on reasoning. According to this Metacognitive dual process account, an initial Type 1 output is accompanied by a `feeling of rightness' (FOR) which regulates the amount of analytic engagement that occurs. When the FOR is persuasive, people are more likely to stay with their original answer, when it is weak, deliberative Type 2 analysis is cued.

The DI dual process theory emphasises the association of beliefs and prior knowledge with Type 1 processes, however, they are also associated with non-logic heuristic processes such as `matching bias'. Empirical support for matching bias comes from a commonly used logical puzzle (since its creation in 1966) known as the Wason Selection Task. The abstract problem presents participants with four cards as illustrated in Figure 1.1:

Conditional rule: If there is an A on one side of the card, then there is a 3 on the other side of the card.

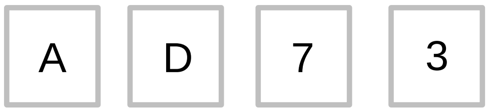


Figure 1.1
Example of the 4 cards presented in the Abstract Selection Task.

Participants are presented with the conditional rule and asked to judge which card they should turn over to determine truth or falsity of the rule. The correct response is A and 7, since the statement can only be falsified by revealing a case where an A card does not have a 3 on the other side. However the task is strongly influenced by the tendency to choose cards that match the cards referred to in the rule (A & 3) (Type 1 process) as opposed to using deliberative logical reasoning (Type 2) to arrive at the correct answer (Evans, 1998). This task has been interpreted as offering experimental support for dual processes when reasoning.

Dual process theories have also received significant support form research on individual differences. For example, Stanovich (1999) has argued that, unlike System 1, System 2 (and therefore Type 2 processing) is associated with measures of general intelligence involved in the ability to decontextualize and decouple reasoning from beliefs. Consistent with this view, abstract deductive reasoning and the ability to resist belief-bias correlates with WM, intelligence and certain cognitive styles (Macpherson & Stanovich, 2007; Stanovich, 2005; Stanovich & West, 1997). People with greater WM capacity or cognitive capacity are more proficient in logical reasoning for both spatial and syllogistic problems (Capon et al., 2003) and specific cases of conditional reasoning (Barrouillet & Lecas, 1999; Markovits, Doyon & Simoneau, 2002). Markovits et al. used concrete and abstract conditional problems and presented participants with three possible conclusions to choose from. They found a consistent correlation between verbal WM and reasoning performance on both types of conditionals. Barrouillet and Lecas (1999) tested children between the ages of 8 and 15 on conditional rules in the form of stories. The children were asked to indicate all the possible examples (or models) in which the rule would be met. The results indicated that the higher the WM capacity the more models the children constructed to finish the task, suggesting that the ability to search for counter examples or alternative models is an analytical process (Verschueren et al., 2005) associated with WM capacity.

Stanovich and West (2008) examined a large number of thinking biases, drawn from the literature on judgments, decision making and reasoning. The biases included the conjunction fallacy, base rate neglect, belief-bias and matching bias. They found that the capacity to resist belief-bias correlated highly with cognitive ability, including matching bias on the four card selection task and belief-bias in syllogisms. Individuals that produced high SAT scores displayed better performance on the selection task and a reduced belief-bias effect. Stanovich and West conjectured that the main source of association between cognitive ability and the biases they examined is down to whether an individual has the capacity for sustained inhibition or the “cognitive decoupling” capacity (also referred to as the ability to separate imaginary situations from real world representations; Stanovich & Toplak, 2012) required to carry out the override of a heuristic response. Take for example a syllogistic reasoning problem where belief and logic conflict: individuals higher in cognitive ability would be better equipped at inhibiting the heuristic cue, based on background knowledge, and holding it in suspension whilst decoupling the structure of the problem from this background knowledge in order to work through the logical structure. Those lower in this ability would be more likely to rely on heuristic Type 1 outputs based on prior knowledge. Stanovich and West (2008) argued that all the tasks they examined which indicated an association with cognitive ability must involve some type of inhibition and/or sustained cognitive decoupling; the central feature of Type 2 processing necessary for hypothetical thinking (Evans, 2007). Similarly, Toplak, West and Stanovich (2011; 2014) showed that cognitive style (as measured by the Cognitive Reflection Task) was a relevant predictor of an individual's tendency to default to heuristic processing (cognitive miserliness). In other words, the better the ability or willingness to engage in Type 2, analytical processes, the better the chance at overriding an incorrect default response.

The view that cognitive ability is related to the capacity to inhibit prepotent responses, based on our beliefs, and decouple them from the logical structure of an argument, is strengthened by work carried out on causal conditional inferences. Evans, Handley, Neilens and Over (2010), instructed participants to reason deductively or pragmatically. In the former case participants were asked whether a conclusion logically followed from a conditional statement assumed to be true. In the latter case participants were instructed to judge their degree of belief in the conclusion supposing the premises, as a way of eliciting belief based reasoning. The findings revealed that those with higher cognitive ability were less belief based in their deductive reasoning but no less so when asked to reason pragmatically. Evans et al. concluded that higher ability participants were better able to decouple their beliefs but only do so when specifically instructed to reason deductively, implying that perhaps pragmatic reasoning is the more natural way of reasoning.

Newstead, et al., (2004) also found that those with higher intellectual ability were less distracted by the believability of a syllogistic argument, more able to decontextualize the problem and respond according to logic rather than belief. Furthermore, higher ability participants were more proficient at handling conflict between content and the logical structure of an argument. This indicates that the ability to abstract rules from their context is an important factor when reasoning. With WM and general intelligence highly correlated (Colom, Abad, Quiroga, Shih & Flores-Mendoza, 2008; Engle, Kane & Tuholski, 1999), it has been suggested that Type 2 processing, unlike Type 1, also engages WM (Evans, 2007; 2009). Research that shows Type 2 thinking, like WM, is susceptible to the effects of aging offer support for this notion, more specifically by the decline in performance on belief-logic conflict problems in syllogistic reasoning tasks as a function of age (Gilinsky & Judd, 1994). De Neys and Van Gelder (2009) corroborated Gilinsky's findings with a life span study on an age range of 12 to 65+ years, showing a curvilinear effect of age on belief-bias, where younger children and the older adults have trouble with reasoning when belief and logic conflict. De Neys et al. concluded that inhibition plays a key role when dealing with conflict in reasoning and decision making (the role of conflict will be discussed further in section 1.4)

In the last 15 years neuro-imaging methods have been increasingly used to investigate the role of different brain areas in reasoning tasks. Goel (2003) demonstrated that the frontal-temporal system is associated with activation of knowledge and experience, whilst the parietal system is linked to formal processing and the visuospatial system for syllogistic reasoning. Goel argues that the former is elementary and effortlessly engaged, corresponding to “heuristic” processing whilst the latter is associated with “universal” processing and requires effort to engage when for example conflict is detected between belief and logic. Further research has suggested that anterior cingulate cortex (ACC) is activated when conflict is detected (also see section 1.4) and the right lateral prefrontal cortex (RLPFC), which is associated with executive control, is required for any subsequent inhibition of intuitive responses, (also see section 1.4.1) more specifically, the right inferior frontal cortex (IFC) is associated with the belief-bias effect (De Neys, Vartanian & Goel, 2008; Goel & Dolan, 2003).

Tsujii and Watanabe (2009) used near-infrared spectroscopy (NIRS) and an attention demanding concurrent task to show that those with enhanced activation in the right IFC perform better on incongruent reasoning trials. This implicates the IFC as having an important role in the inhibition of default responses, explaining why a secondary load that impairs IFC activation leads to increases in belief-bias. This research suggests that distinct brain areas are required for belief based and logic based reasoning, which is consistent with the DI dual process account (Evans & Stanovich, 2013). Finally, research involving the stimulation to the right inferior frontal cortex using transcranial magnetic stimulation (TMS) has been shown to enhance the belief-bias effect, whilst stimulation to the left inferior frontal cortex reduces belief-bias. Tsujii, Masuda, Akiyama and Watanabe (2010) argue that semantic information processing interferes with reasoning performance on incongruent trials, therefore when the right IFC is stimulated and inhibition is affected subjects fail to inhibit semantic processing and belief-bias is enhanced. In contrast, stimulation to the left IFC, improved reasoning performance on conflict trials by blocking irrelevant belief based responses.

This section has provided evidence, across various fields of research, for dual process theories. More specifically the research has shown that responses based on prior knowledge are activated by default and complete quickly, whilst reasoning based on the logical structure of a problem requires effortful, Type 2 processing, in support of the DI account. This review has also highlighted that Type 2 processing relies on WM resources, is linked to cognitive style and ability and develops with age. Furthermore, neuropsychological evidence suggests that belief and logic based reasoning are neurologically dissociable. However, despite the seemingly large body of evidence in support of a distinction between processing types, a number of authors have suggested that the data can be equally explained through a single system account. We consider their critiques in detail in the next section.

1.3.1 Critique of Dual Processing

As discussed in the previous section DP theorists make a distinction between Type 1 and Type 2 processes. The former often characterised as implicit, quick and effortless whilst the latter is often categorised as explicit, slow and demanding on WM resources (Evans, 2008). However several authors argue that there are a multiplicity of dual processes beyond reasoning, such as social cognition and learning and that there is no consistency in terms of the characteristics attributed to these different types of processing. Newstead (2000) makes reference to the conflicting views on the characteristics of Type 2 thinking, for example; some research has made reference to its association with intelligence (Evans, 2010a) where as others have found few correlations between intelligence and rational processing (Klaczynski, Gordon & Fauth, 1997).

Gigerenzer and Regier (1996) criticise the terminology and categorisation of the characteristics involved in the two systems proposed by Sloman (1996); for example, how the representativeness heuristic (part of the associative system) can be expressed as "likelihood" in the rule-based language of probability. They suggest that the occurrence of two opposing responses in the case of the Wason Selection Task (Wason, 1966;

1968) or Conjunction Fallacy (Tversky & Kahneman, 1983) does not necessarily support two systems and could instead be the result of one formal rule based system, resulting in linguistic uncertainty, conjecture or competition between opposing formal rules (Gigerenzer & Regier, 1996). Gigerenzer (1996) argues that heuristics such as representativeness or availability are vague in terms of their definition and do not provide adequate explanation of the underlying cognitive processes involved in generating a response to judgments. Gigerenzer and Goldstein's (1996) carried out research on inferential tasks where choices had to be made between two alternatives, based on knowledge retrieved from memory. Using computer based simulations of people with varying degrees of knowledge; they questioned how well a cognitive algorithm would perform in a real-world environment. Gigerenzer et al. found that certain `frugal' algorithms could draw as many correct inferences about a real world environment as algorithms representing rational (logical) inference. In other words, Gigerenzer's research showed that judgments based on knowledge and shortcuts do not necessarily result in error, as previous accounts claim, and that the fast and frugal heuristics acquired allow for the delivery of optimal solutions to real world problems.

Osman (2002) used variations of the Wason Selection Task to investigate some of the claims made about the errors commonly associated with the task (Wason & Evans, 1975). The research focused on the assertion that unconscious biases encourage participants to focus their attention on cards matching the conditional rule. Osman found no sign of this attentional bias on the latency data and also revealed that tutoring on the task actually improved performance, refuting the inflexibility of unconscious processing that previous accounts claim. She argues that the results highlight the lack of sufficient evidence for unconscious reasoning and fail to support dual process claims made about the selection task.

Generally Osman (2002) advises that the techniques used to expose unconscious reasoning are in need of improvement. More specifically Osman (2004) proposes that the evidence used to support separate systems is applicable to, and more consistent with a single system account. Osman (2004) introduces the dynamic graded continuum (DGC) model (Cleeremans and Jiménez, 2002) as a reasonable single system alternative that makes an important distinction between implicit and automatic reasoning. The DGC model proposes about the quality of a `representation' in the mind changes (i.e. its strength or distinctiveness) along the continuum from implicit to explicit to automatic. Consciousness is seen as graded and dynamic according to subjective experience, for example; changing gears when learning to drive starts with explicit instruction that through repetition of the behaviour gradually moves along the continuum until it reaches automaticity, a behaviour that acts outside of conscious control. The same process can be used to explain the progression through reasoning, assigning different roles to consciousness when reasoning rather than different systems underlying responses. Similarly she interprets individual differences in reasoning as relating to differences in the degree rather than the kind of reasoning used.

Keren and Schul (2009) also propose a uni-model, motivated by what they describe as a lack of clarity with dual-system theories which have been constructed on problematic methodologies and insufficient pragmatic evidence. For example, the two-system model has been used to explain conflict among mental states but Keren and Schul argue that the presence of conflict is not adequate evidence for two independent systems. Consider for example the belief-bias effect which attributes belief based reasoning to System 1 and logical inferences to System 2 processes. Evans & Curtis-Holmes (2005) showed that System 1 predominates under limited time and increases belief-bias as evidence for dual processing. Keren and Schul argue that the effect can be explained by the logical validity and believability being two distinct types of external criteria that the single system can be used to evaluate. Similarly with the `Linda problem' (Tversky & Kahneman, 1983) Keren and Schul suggest that rather than the individual considering two conflicting beliefs simultaneously, perhaps when they reason about the combined option in the Linda problem (Linda is a bank teller and a feminist) they temporarily forget the single option (Linda is a bank teller). Therefore, rather than a dichotomy of sequential or parallel processing, maybe the level or awareness of an alternative response fluctuates continuously when reasoning. In other words a single system can shift between many mental states to solve different tasks and these states are defined by an assembly of different features such as speed, level of control, awareness, etc… these features join in different ways at different times depending on the goal and also environmental limitations.

Finally, Kruglanski and Gigerenzer (2011) propose a unified theory based on rule processing and present a number of pragmatic arguments for this theory describing intuitive and deliberate judgments as rule-based. The rules can be optimising or heuristic and both types of judgment can use the same rules. For example, the recognition heuristic used to acknowledge a name or face is not deliberative but it can be used intentionally as a strategy, i.e. choosing a horse in at the Grand National. Kruglanski and Gigerenzer suggest that rules are based on cognitive capacity in that the type of rule selected or the speed and accuracy at which a rule is applied will depend on individual differences in these cognitive capacities. They argue that both heuristic and deliberative rules can be equally difficult or easy to apply but there is a relationship between an individual's processing potential and rule application. Under this unified model belief-bias is understood as conflict between rules, suggesting that improved deduction through tutoring can be explained by improved focal rule application. Kruglanski and Gigerenzer's propose a two-step rule selection process with this model, whereby memory content and the task limit the rules available to an individual, whilst processing potential and environmental factors regulate the final rule selection.

Aside from the empirical support already offered for dual processing (section 1.3), Evans and Stanovich (2013) offer some additional points for consideration in respect to the criticisms extended. Firstly they argue that Kruglanksi and Gigerenzer (2011) overlook the evidence that supports the case for qualitatively separate processing types and secondly, they contend that placing all dual process/system theories under the same umbrella is problematic. For example, Type 1 and Type 2 processing (Evans, 2009) is different to distinct modes of processing, which relate to different cognitive styles applied in Type 2 thinking. Modes of thinking can rely on personality, culture, situation and motivation and can vary continuously. Type differences relate to cognitive ability and modes relate to cognitive styles such as Need for Cognition (Cacioppo & Petty, 1982) and Active Open Minded Thinking (Stanovich & West, 1997; 2008).

Evans and Stanovich (2013) dispute the claim that intuition and deliberation being rule-based is enough to support or refute the idea that they come from distinct cognitive systems. Moreover, they do not dispute the possibility that Type 1 associative processing can be formed through the implementation of rules. Evans (2010a) offers the Two Mind theory as a replacement for the Two System theory that has come under some criticism. The two minds consist of the *old* intuitive mind and the *new* reflective mind, both of which comprise a multitude of systems. The difference is that the new Two Mind theory implies an autonomous old mind that can influence behaviour directly whilst both Type 1 and Type 2 processing are part of the new mind and Type 2 processes must cooperate with working memory.

Furthermore, Evans (2010b) suggests that the distinction between intuitive and heuristics processes has been somewhat confused throughout the literature. He suggests that “Intuitive” should be considered a Type 1 process and “Heuristic” as Type 2, because heuristics are based on simple yet explicit rules but can be as quick as a Type 1 process. Finally, Stanovich (2009a; 2011) offers a detailed extension to his original model, implying hierarchies of control rather than separate systems. The model referred to as the “Tripartite” model suggests that the autonomous mind houses Type 1 processes whereas the algorithmic mind related to cognitive ability, contains the algorithmic level of Type 2 processing and the reflective mind, linked to epistemic dispositions, houses the reflective level of Type 2 processing. The autonomous system can be inhibited by the algorithmic mind but the override is initiated by higher level control of the reflective mind (Evans, 2011).

In sum, Evans and Stanovich (2013) agree that it is ill-advised to assume all the suggested characteristics of the two processing types, are necessary and defining. However, they do argue for a clear distinction between Type 1 and Type 2 (terminology we employ throughout this thesis), where Type 1 processes are defined by their autonomy and Type 2 processes are identified as drawing on the working memory and allow for hypothetical thought through the process of “cognitive decoupling” (Stanovich, 2009b; 2011). Additionally, those with higher WM capacity will be better able to inhibit autonomous responses and engage in Type 2 processing as opposed to relying on automatic Type 1 responses (Stanovich & West, 1998a).

Although DPT has come under considerable criticism and these critiques have been defended in a number of ways by Evans and Stanovich as illustrated above; there is also considerable empirical support for the account, as shown in section 1.3. To reiterate, default processing focuses heavily on how our prior beliefs are the intuitive output that bias our decision making processes (Evans, 2006), whilst logical reasoning necessitates effortful Type 2 processes. However, recent research has proposed that reasoning on the basis of beliefs may not be as fast and automatic as these accounts claim (Handley et al., 2011). In the next section we begin by reviewing research demonstrating that reasoners can detect when there is conflict between responses. This suggests that responses can be triggered simultaneously rather than serially as most DI DP models claim. In addition to this we review the research indicating that reasoning based upon logical structure can often be completed quickly and intuitively.

1.4 Conflict Detection and Logical Intuition

As mentioned previously, the traditional DI account defines reasoning on the basis of knowledge or belief as a Type 1 process triggered by default, whilst a response based upon the logical structure of a problem draws on Type 2 processes. This section reviews some recent evidence that is inconsistent with this claim by suggesting that some structural processing resembles the fast and effortless characteristics of a Type 1 response. We start with research showing that people have the ability to detect conflict between competing responses, which is contrary to the claims made by the DI DPT.

De Neys and Glumicic (2008) examined the responses to base rate problems where the description and the base rate where either congruent (no-conflict), incongruent (in conflict) or neutral (neutral description). They demonstrated that participants took longer to process the conflict problems but verbal protocols showed that they were explicitly unware of any conflict. Furthermore, even when participants gave a response based upon the description rather than the base rate, evidence suggested that (through retrieval of base rate information) the base rates had also been thoroughly processed. De Neys and Glumicic claimed that implicit conflict detection, not only suggests intuitive awareness of the base rates, but is consistent with a parallel model. Stupple and Ball (2008) also offer evidence of conflict detection using inspection-time analysis on syllogisms to show that conflict problems, where logic and belief are inconsistent, lead to increased processing latencies relative to non-conflict problems (also see Ball, Philips, Wade and Quayle, 2006). They offer an explanation which assumes that concurrent outputs from both heuristic and analytic processes are available, with the latter requiring more time to resolve.

These studies support the notion that conflict is intuitively detected outside the requirement for Type 2 processing. De Neys, Moyens & Vansteenwegen's, (2010) research also confirmed the effortlessness of conflict detection present with syllogisms. They used skin conductance responses (SCRs) to show autonomic arousal with inconsistent conflict problems implying a "gut" feeling of incorrectness. Gangemi, Bourgeois-Gironde & Mancini (2014) refer to this gut feeling as Feeling of Error (FOE), explained as the detection of the conflict between an intuitively incorrect and normatively correct answer to a problem. Gangemi et al. used the bat and ball problem (Frederick, 2005) and a non-conflict version of the problem to examine the FOE on both generation and evaluation versions of the task. They followed these with a FOE questionnaire to assess the level of cognitive discomfort experienced from completing the problems. Gangemi et al. found that those who failed at the task experienced more FOE, suggesting that it is a reliable signal that errors are present and assume that it results from the detection of conflict impacting on reasoning performance. Consequently, it appears that even if detection is not explicitly expressed (e.g., De Neys & Glumicic, 2008) and judgments are often biased, people do appear to sense their inaccuracies.

De Neys, Vartanian & Goel (2008) introduced a neuro-scientific approach which analysed fMRI data captured on the classic "lawyer-engineer" base rate problem. Similar to De Neys and Glumicic (2008) they used congruent, incongruent, neutral (description neutral) and heuristic-control (base rates neutral) base rate problems. Their results showed that the area of the brain involved in conflict detection was activated even when participants responded based upon the stereotype, suggesting that they intuitively detected their bias. In other words, even though participants gave a stereotyped response, on some level they were aware that it conflicted with the base rate. If people were not intuitively sensitive to the statistically correct response then the conflict observed in the anterior singular cortex would not have been detected (De Neys & Franssens, 2007; De Neys et al., 2008).

According to the DI model, Type 2 processing does not start until conflict between Type 1 and Type 2 is detected; the question this model fails to address is how conflict can be detected in the first place if Type 2 processing has not been initiated? Bank and Hope (2014) conducted a study on belief-bias in relational reasoning using valid, determinately invalid and indeterminately invalid problems, each with both believable and unbelievable conclusions. They monitored EEG activity whilst participants solved the problems and found greater P3 amplitude (determined by the allocation of attentional resources when WM is updated) was present when belief and logic conflicted. Furthermore, P3 latencies were equivalent for belief and logic responses, indicating that they both influence reasoning at the same time and relatively early. These findings fall in favour of a Parallel Competitive model (PC) (Epstein, 1994; Handley et al., 2011; Sloman, 1996) consistent with the notion that logic (structural) and belief (relevant knowledge) are processed simultaneously as opposed to serially: we return to the PC model in the next section.

The studies on conflict detection and the idea that a logical response can be triggered early and at the same time as a belief response suggests that people are sensitive to logical structure, which may indicate a degree of logical intuition. This interpretation is supported by Radar & Sloutsky (2002) who presented participants with the conditional and second premise to Modus Ponens (MP) problems within a story context, for example; *if the weather is nice then Ed takes a walk; the weather was nice*; and participants had to decide if a particular word had appeared within the story. The findings indicated that participants were more likely to falsely identify that the conclusion had been presented, suggesting that MP were drawn implicitly at the comprehension stage, which is a rapid, automatic Type 1 process. This also extended to problems based on Affirmation of the Consequent. Similarly, Leo & Greene (2008) conducted research on inferences made about the relationship between the novel configurations of an array of faces presented, whereby participants were unaware of the task requirements. They found that inferences can be drawn without deliberate strategies and that logical relational processing can take place irrespective of task awareness.

Morsanyi and Handley (2012) also suggested that a degree of logicality can be activated intuitively, with the use of syllogisms and liking ratings. They measured intuitive contributions to reasoning by examining how much participants `liked' the concluding statement to an argument and found an effect of logical validity on liking judgments. Liking ratings are based on intuitive processing and are said to reflect a person's affective state. With logicality affecting liking ratings, Morsanyi and Handley suggested that the logic effect was elicited by the conceptual fluency of processing the conclusion. In addition, they found that when emotional responses were attributed to background music the logic effect reduced, confirming that the participant's affective state induced the effect. Indeed the logic effect remained present after reducing premise presentation time which confirmed that logical analysis does not always rely on time consuming, effortful processing (see Singmann, Klauer & Kellen, 2014, for a recent critique of this research).

Of course we cannot forget the ample research on the belief-bias (see section 1.2.2) promoting the effortful nature of structural processing, but we should consider the way in which participants are instructed to carry out the reasoning task. In the majority of bias studies participants are specifically directed to judge whether a conclusion logically follows from the premises presented. Therefore, perhaps belief-bias can be explained in terms of an individual's inability to adhere to logical instruction. For example; Evans et al., (2010) showed that those higher in cognitive ability were more proficient at decoupling their beliefs but only when specifically instructed to do so. This might suggest that the demanding nature of logic based reasoning may actually be down to the effort required to apply a specific set of complex instructions.

Handley, Newstead and Trippas (2011) introduced a new instructional set to their research on conditional reasoning which required participants to judge the believability of a conclusion as well as its logical validity. They examined the processing accuracy and latency data on both conflict and no-conflict problems. Depending on whether participants were instructed to reasoning logically or judge the believability of a conclusion, conflict problems would prompt different responses; whereas no-conflict problems elicited the same response regardless of instruction. They found that logical inferences on both MP and disjunctive arguments are accomplished relatively automatically and are immune to the influence of beliefs. Judging conclusion believability, on the other hand, is a slower process that is subject to interference from logical analysis. This effect was present in both a between and within subjects manipulation and extended to both accuracy and latency data. Their findings contradict the predictions made based on a DI account of reasoning and what would be expected if belief based judgments were based on fast, automatic Type 1 processes and logic based judgments were based on slow, deliberative Type 2 processes. Pennycook, Trippas, Handley & Thompson (2013) extended these findings to base rate problems whereby participants were instructed to respond on the basis of the base rate information or the personality description embedded in the problem. The former would require participants to disregard any background knowledge whilst the latter would require participants to overlook the statistical base rate information. They found that both statistical and stereotyped information interfered with one another and produced similar processing times, suggesting that information based on either the base rate or background knowledge was triggered automatically and in parallel.

Once again, both studies demonstrate that judgments based on the logical structure of the argument can be accomplished quickly and intuitively, and contrary to the DI account, they imply that belief judgments require some effortful processing. Other research has also emphasised the demanding nature of belief based processing, for example; belief-bias on conditional inferences has been shown to be unaffected by speeded problem presentation unlike syllogistic reasoning (Evans, Handley & Bacon, 2009). Evans et al. (2009) found that the impact of speeded presentation was equal under deductive and pragmatic instruction. Newstead et al., (2004) found a strong relationship between cognitive ability (as measured by the AH5) and deontic selection tasks, which depend upon accessing relevant knowledge. Furthermore, Experientiality (as measured by REI, see Pacini & Epstein, 1999) as a measure of thinking disposition, did not correlate with belief-bias or the deontic selection task and cognitive motivation was a poor predictor of logical performance. Additionally, recent research has shown that among pre-adolescent children, biased responding on reasoning tasks is more, rather than less, common in participants who score higher on measures of working memory (Morsanyi & Handley, 2008).

Recent research has also shown that working memory load can result in a decrease in pragmatic responses on reasoning tasks accompanied by an increase in logical ones. For example, De Neys and Schaeken (2007) presented participants with sentence verification tasks, which involved awkward and under informative sentences that prompted scalar implicatures; for example the sentence *`some oaks are trees'* would cue the implicature *`not all oaks are trees'*. The sentence is logically true but the cued implicature is false compared to general knowledge about trees. Therefore when asked to judge whether the sentence is true or false the implicature leads participants to incorrectly judge the sentence as false. Some authors have assumed that implicatures are generated automatically (Levinson, 2000) however, De Neys and Schaeken (2007) showed that certain pragmatic implicatures were reduced under cognitive load whilst logical performance increased, suggesting that judgments based on knowledge must also rely upon Type 2 processes.

Finally, De Neys, Schaeken & d'Ydewalle (2005) presented participants with causal conditional statements such as *"if the air conditioning is turned on, then you feel cold"*, followed by factual statements like *"you feel cold, but the air conditioning was not turned on"* and asked them to generate counterexamples to the rule. The results revealed that searching for counterexample information draws on WM capacity and the efficacy with which they are retrieved is reduced by WM load. Again these findings suggest that inferences based on prior knowledge can influence reasoning in ways that are slow, reflective and dependant on Type 2 processing (also see Verschueren et al., 2005).

In this section we discussed conflict detection and how it supports the concept of logical intuition, suggesting that judgments based on the structure of an argument can be accomplished relatively automatically. The literature also revealed that when instructed to reason on the basis of belief, judgments are slow, more prone to error and are impeded by the validity of the argument. Furthermore, the reliance on background knowledge to reason appears to increase as WM capacity develops (Morsanyi & Handley, 2008), all of which corroborate the view that judgments based upon beliefs can require effortful processing. We also briefly mentioned how the findings presented in this section are in conflict with the DI dual processing account and are more consistent with belief based and logical processing operating in parallel. In the next section we will discuss the PC account in more detail and the role of inhibition in relation to the model.

1.4.1 Parallel Processing and Inhibition

Handley et al. (2011) suggest that reasoning about beliefs is often demanding and relies on the integration of relevant knowledge with aspects of problem structure in order to generate a novel response. As mentioned previously (see section1.3.1) an alternate dual process model is the Parallel Competitive (PC) model (Epstein, 1994; Handley et al., 2011; Sloman, 1996) which is common in social psychology (Chaiken & Trope, 1999) and has been applied to the judgment and decision making literature (Sloman, 1996). According to Sloman's (1996) interpretation of the PC model, belief and logic influence each other bi-directionally, suggesting that people always simultaneously engage in both Type 1 (heuristic) and Type 2 (analytic) processing which occur in parallel and require conflict resolution through the inhibition of the inappropriate response. This model is more consistent with the findings in the previous section and more importantly it offers an explanation for the experience of conflict detection when reasoning, which the DI model fails to do.

The PC account suggests that for an analytical Type 2 response to complete, it must inhibit the heuristic Type 1 output and the dominance of heuristic responding as demonstrated by the literature on biases, is the result of inhibition failure; for example, Houdé et al. (2000) monitored brain activation to show that increasing inhibitory processing through training, improved selection rate on the Wason Selection Task (Wason, 1966; 1968), suggesting that poor performance before inhibition training was the result of poor inhibition of the intuitive matching response rather than a lack of logical knowledge. Additionally, Moutier & Houdé (2003) confirmed that inhibition training was better at improving reasoning performance on classical conjunction bias, than logical training.

De Neys, Novitskiy, Ramautar & Wagemans (2010) conducted a study on both base rate and conjunction problems where the stereotypical description cued a heuristic response that conflicted with the `normatively' correct answer. The results from both problem sets created two groups of `most' and `least' biased reasoners which were then examined using an EEG study of executive monitoring and inhibition skills. The results indicated that even though most reasoners detect they are biased, those that perform better on the reasoning problems are characterised by a superior inhibition capacity. This emphasises the key role of inhibition in overriding tempting yet erroneous intuitions to conflict problems.

Markovits and Barrouillet (2002) suggested that an essential component in the development of conditional reasoning relies on an individual's ability to inhibit information that does not fit with the presented problem. Simoneau and Markovits' (2003) research lent further support, implying that reasoners can reason logically when given specific instruction suggesting that the ability to "inhibit" counterfactual information can explain why some people have the aptitude to "suppose premises are true" when they are not actually true. They also suggested that selective inhibitory processes underlie the development of any associative system which allows increasingly complex processing without the help of any explicit rule.

De Neys & Franssens (2009) examined the inhibition process in reasoning with syllogisms, using a lexical decision task to test whether supressing beliefs during reasoning would affect their subsequent recall. Results showed that conflict between belief and logic increased latencies on lexical decisions for target words, and that the retrieval of these believable words was impaired. They concluded that this was not due to a failure to detect conflict or the requirement to inhibit the heuristic response but rather a failure to complete the inhibition process, in favour of a parallel processing approach when reasoning.

However there has been debate over the efficiency of a fully parallel model, for example De Neys & Glumicic (2008) argue that such a model dismisses the advantages of a heuristic route; in other words, if the heuristic process arrives at the correct response, why would people redundantly complete the analytical process? They hypothesised that if people complete the analytical process in all circumstances then base-rates on congruent and incongruent problems should be attended to equally. However, their research did not support this hypothesis or a purely parallel model. Equally, the research on conflict detection makes a solely serial model highly problematic; therefore De Neys (2012) offers an alternative parallel model. He proposes parallel activation of two distinctive intuitive responses; one logical and one heuristic. The heuristic response will call on stereotypical associations and semantics whilst the logical response will pull on typically logical and probabilistic rules. If and when the responses conflict, then the deliberative and effortful processing is engaged. In order to explain why the heuristic system often dominates, he suggests that activation levels amongst the two intuitive responses differ. Where the heuristic response produces a more salient activation (De Neys, 2014) it nevertheless produces doubt and can account for the conflict detection (De Neys & Glumicic, 2008) as well as explaining the need for inhibition of the heuristic response. However, when the heuristic response is so persuasive people may not engage the inhibition process, alternatively some people may be aware of the conflict, try and block the compelling heuristic response but fail to do so, which is referred to as the "flawless detection view" (De Neys, Cromheeke, & Osman, 2011). Again this supports the idea that people can be "intuitive logicians", influenced by traditional logical principles when making decisions.

1.5 Summary

In this chapter we introduced the literature on DPT in reasoning, concentrating on the broadly accepted DI approach (Evans, 2008) which associates belief based reasoning with default, Type 1 processes and reasoning based on the logical structure of an argument as pulling on Type 2 processes. Research on the belief-bias effect (Evans, Barston & Pollard, 1983) offers considerable support for the DI account, showing how the effect can be increased with time constraints (Evans & Curtis-Holmes, 2005) and under cognitive load (De Neys, 2006; Quayle & Ball, 2000) and how it can be reduced when participants are emotionally charged (Goel & Vartanian, 2011) or through explicit instructions to reason logically (Evans, Newstead, Allen, & Pollard, 1994). However, the evidence of early conflict detection implies that people are implicitly aware of two competing sources of information based on knowledge (beliefs) and problem structure (validity) which offers support for an alternative PC dual process model of reasoning. Further research proposes the idea of intuitive logic (De Neys 2012; 2014) that has the ability to interfere with the believability of a conclusion (Handley et al., 2011). The concept of the intuitive logician provides a more optimistic view of human reasoning, in that it suggests that people have an intuitive sense of logical structure.

1.6 Rationale and Structure

The evidence supporting the intuitive nature of logic based reasoning (De Neys 2012; Morsanyi & Handley, 2012) and the view that belief based judgments also require effortful processing (Morsanyi & Handley, 2008; Verschueren et al., 2005) provides the framework for the research presented in this thesis.

Our broad objective is to investigate whether belief based judgments are in fact resource demanding, using Handley et al.'s (2011) novel methodology as our foundation. As we have seen, they investigated the accuracy and time course of processing when participants were instructed to respond, either, on the basis of logical validity or the believability of a conclusion. Their findings suggested that a) certain logical inferences are accomplished relatively automatically and are immune to the influence of beliefs and b) judging conclusion believability appears to be a slower process that is subject to interference from logical analysis. This research challenges the DI account of reasoning which is contrary to what would be expected if belief based judgments were based on fast, automatic Type 1 processes and logic based judgments were based on slow, deliberate Type 2 processes.

Substantiated by their research and the premise that belief based judgments require effortful Type 2 processing, we investigate whether belief judgments will be equally, if not more so, affected by a secondary task as logical judgments. Each experiment throughout this thesis implements the original methodology of Handley et al. (2011) which allows us to evaluate the degree to which logic and beliefs interfere with one another. In Chapters 2 and 3, the interference between belief and logic is examined whilst participants engage in a secondary task demanding of working memory resources. Chapter 2 examines the effect of a Dot Memory task and Random Number Generation (RNG) on conditional Modus Ponens (MP) arguments. Chapter 3 introduces more complex disjunctive arguments and several individual differences measures for investigation, as well as examining performance on the secondary task.

An additional aim of the current research is to investigate the underlying executive processes necessary for belief and logic based reasoning; essentially determining whether their demand on executive resources differentiates. In Chapter 4 we investigate the role of working memory with the use of two original experimental designs combining a Stroop task (Stroop, 1935) and an N-Back task (Kirchner, 1958) with simple (MP) and complex (Disjunctives) reasoning questions. Finally, in Chapter 5 we discuss our findings in relation the dual process framework, followed by recommendations for future research.

Chapter 2: The effect of secondary load on simple judgments.

2.1 Introduction

In Chapter 1 we discussed the *Default Interventionists* (DI) interpretation of belief-bias whereby belief responses are considered Type 1 processes that are generated by default and may require intervention from more deliberate Type 2 processes at times (Evans, 2008) (see Figure 2.1 for representation).

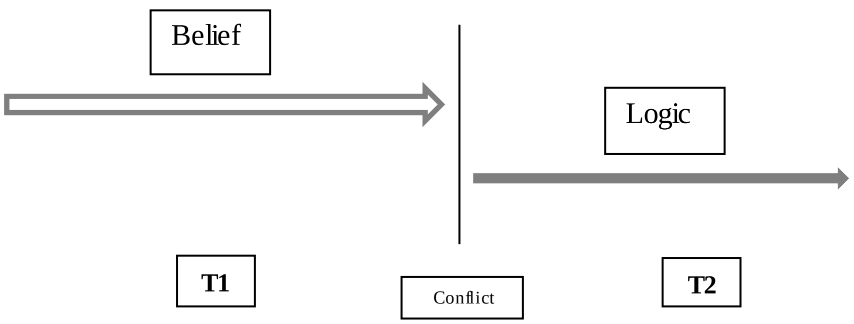


Figure 2.1
An illustration of the DI model shows that a Belief based response, to a simple argument, is a Type 1, intuitive process that is triggered by default and completes early. A Logical response on the other hand, requires deliberate Type 2 processing and takes longer to complete. The vertical line represents the transition between T1 and T2 processing and when conflict is implicitly detected.

Default processing tends be associated with how prior beliefs and the use of heuristics impede on our decision making processes (Evans, 2006). In contrast other research has shown that belief based judgments may not be as fast and automatic as previous accounts claim (Evans, Handley & Bacon, 2009; Morsanyi & Handley, 2008; Newstead, Handley, Harley, Wright & Farrelly, 2004). In fact it is becoming evident that some belief based judgments are effortful, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response. For example, De Neys, Schaeken & d'Ydewalle (2005), demonstrated with causal conditional reasoning, that the retrieval and integration of knowledge from counterexamples was directly associated with Working Memory (WM) capacity. Furthermore, Verschueren, Schaeken & d'Ydewalle (2005) highlighted the slow analytic nature of processing counterexample information and the ability this information has to override conclusions based on probabilistic processing.

In the previous chapter we introduced Handley, Newstead & Trippas (2011) novel methodology that required participants to evaluate the conclusions of logical arguments on the basis of either logical validity or believability. They consistently found that belief based judgments took significantly longer than those made under logical instructions and judgments relating to strongly held beliefs could themselves be undermined if they were inconsistent with the logical structure of an argument. This is contrary to predictions made by the DI account and inconsistent with the idea that belief based judgments are based on automatic Type 1 processes and logic judgments require effortful, Type 2 processes.

Handley et al. (2011) proposed an alternative Parallel Competitive (PC) model to interpret their findings (see Figure 2.2).

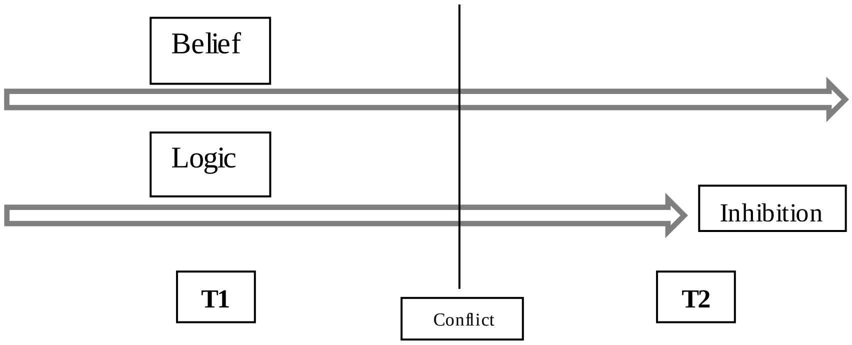


Figure 2.2
An illustration of the PC model shows that both Belief and Logic based responses to a simple argument are initiated in parallel. Both responses require Type 2 processing, but Logic based responses complete first and require inhibiting in order for a Belief based response to complete.

They suggested that the reasoning problems they presented prompted a rapid logical response (as opposed to a default belief based response) which required inhibition when presented with a belief based instructional set. Therefore, if responding on the basis of beliefs requires effortful processing and the inhibition of conflicting responses (Evans, 2003), then based on Handley et al. belief based judgments should take longer and be more affected by a secondary load than logical judgments.

The aim of the current chapter is to evaluate the degree to which logic and belief based judgments interfere with one another whilst participants engage in a secondary task. A secondary task that burdens WM resources will tax analytic, Type 2 processing and the ability to override heuristic, Type 1 outputs (De Neys, 2006). Consequently, judgments that require effortful Type 2 processing will be impeded by a demanding secondary task. This chapter explores the level of effortful processing required for belief and logic based judgments in line with the PC model (Figure 2.2). In accordance with Handley et al. belief and logic based responses are prompted simultaneously and logical responses complete first; if they are correct in their findings then we predict that when instructed to judge the believability of a conclusion, whilst engaged in a secondary task, fast completing logical responses will become harder to inhibit and thus hinder more effortful, belief based judgments.

De Neys (2006) employed the secondary load method on syllogistic reasoning with the use of a spatial storage task and showed that belief-bias increased with a secondary load. De Neys interpreted this finding as indicating the heuristic nature of belief based reasoning and the notion that erroneous reasoning in the case of a belief-logic conflict is not only associated with, but also directly caused by, limitations in executive resources. However, like most research on belief-bias in reasoning, participants were only instructed to determine whether the conclusion followed logically from the premises of the syllogism. Consequently, the evidence of increased belief-bias under load may not be due to a reduced capacity to inhibit Type 1 processing but rather the result of participants failure to appropriately apply the correct instructions when their WM is impacted on.

Handley et al. (2011) proposed that their evidence *supported* the role of Type 2 processes in belief based judgments. Therefore, we will employ Handley et al.'s original methodology, that asked participants to judge conclusion validity as well as conclusion *believability*, whilst engaged in De Neys' (2006) `Dot Memory Task' as a direct test of the DI account. We hypothesise that when instructed to judge the believability of an argument, participants will be more affected by the secondary load than when instructed to judge conclusion validity.

2.2 Experiment 1

The overall aim of Experiment 1 is to determine whether applying a secondary load will have a differential impact on belief and logic judgments for simple Modus Ponens (MP) arguments, specifically when the logical validity and belief in the conclusion are in conflict (see Table 2.1). According to the DI model (see Figure 2.1) that highlights the default nature of belief based responses; logic judgments should be more affected by a secondary load. Alternatively, and in accordance with the PC model (see Figure 2.2), if belief based judgments are demanding on Type 2 processing, specifically the WM resources required to inhibit logic based inferences, then they should be more affected by a demanding dot memory task (see Figure 2.3).

2.2.1 Method

Participants

A total of 81 psychology undergraduates from Plymouth University, England, took part in Experiment 1, in return for a course credit. Participants consisted of 67 female and 14 male students.

Design & Material

A 2 x 2 x 2 mixed design was used where participants were randomly assigned to the high load or low load group. Each participant received the same 64 reasoning problems under both belief and logic instructions (Instruction type). Half the problems were conflict problems (with conclusions that were valid and unbelievable or invalid and believable) and the other half were no-conflict problems (with conclusions that were valid and believable or invalid and unbelievable). Refer to the reasoning task below for more details.

Reasoning Task:

Table 2.1
Examples of the different types of Modus Ponens (MP) arguments used in Experiments 1 – 3.

Conflict Arguments	**No-conflict Arguments**
Argument 1 If it is raining then the street is dry It is raining (A) The street is dry Logic: ✓ Belief: X	*Argument 2* If it is raining then the street is wet It is raining (C) The street is wet Logic: ✓ Belief: ✓
Argument 3 If it is raining then the street is dry It is raining (B) The street is wet Logic: X Belief: ✓	*Argument 4* If it is raining then the street is wet It is raining (D) The street is dry Logic: X Belief: X

The reasoning task was presented to each participant on a computer screen. Participants received 64 conditional reasoning problems in Modus Ponens format (see Appendix B for a complete set of the stimuli). Thirty two were under belief instruction and 32 under logical instruction. In each of these sets there were 16 conflict trials, 16 no-conflict trials and in each case eight required a yes response or a press of the `s' key on the key board and eight which required a press of the `k' key for a negative response. Table 2.1 demonstrates the type of questions presented to each participant. Argument 1 shows that the conclusion of the argument logically follows from the premise but is in conflict with what we believe to be true about the world. Argument 3 shows a problem where the conclusion is believable but does not follow logically from the premises. For arguments 2 and 4 both the form (logical validity) and content (if it is believable) lead to the same conclusion, hence there is no-conflict. The four argument types were randomly presented throughout the experiment with an equal number of belief and logic based responses.

Table 2.2
Examples of how an argument was presented under both instructional sets.

Conditional argument presented under Logical Instruction	**Conditional argument presented under Belief Instruction**
If a dog is barking then it is silent (Premise 1)	If a dog is barking then it is silent (Premise 1)
Suppose a dog is barking (Premise 2)	Suppose a dog is barking (Premise 2)
Does it follow that the dog silent?	Does it follow that dog silent?
Valid ? or Invalid?	Believable? or Unbelievable?

For each problem the first (major) premise was presented alone; then when the spacebar was pressed the first premise disappeared and the second (categorical) premise, conclusion and two answers to choose from appeared on the screen. Before the experiment started participants received eight practice trials with feedback, which covered all four argument types (see Table 2.1). Four were presented alone and four were presented with a dot matrix pattern to memorise. The dot pattern was either simple or complex and was presented before the MP argument. When participants had responded to the argument, an empty matrix grid appeared where they were required to replicate the pattern they had memorised earlier (see Dot Matrix Memory Task).

Dot Matrix Memory Task:

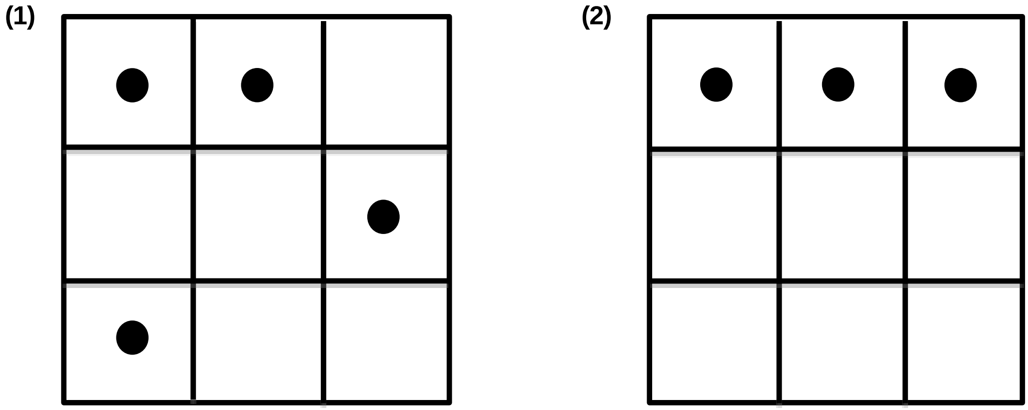


Figure 2.3
A Dot Pattern example in the High Load (1) and Low Load (2) Conditions.

Participants were asked to memorize a 3x3 matrix filled with three or four dots, which was presented before each reasoning question for 850ms (De Neys, 2006). They were then required to reproduce the dot pattern after completing each reasoning task. Participants were split into high and low load conditions. The high load consisted of a matrix filled with a complex four-dot pattern, i.e., a ‘‘two- or three-piece’’ pattern based on Bethell-Fox and Shepard (1988), and Verschueren, Schaeken and d'Ydewalle (2004). The two piece pattern meant that two or three of the dots that were adjacent to each other could not be adjacent to the remaining dot(s), (but they could connect diagonally) making up a two piece pattern. The three piece pattern meant that two connecting dots could not be adjacent to the remaining two dots, which could also not rest adjacent to each other (but could touch diagonally) making up a three piece pattern. The low load condition had a three dot pattern in a horizontal line (i.e., a ‘‘one-piece’’ pattern in Bethell-Fox & Shepard's (1988) terms). The low load was to ensure a minimal burden was placed on executive resources. One hundred and twenty eight matrices were created altogether, 64 with a complex pattern, 36 of which were made up of two pieces and 28 three pieces, the remaining 64 were made up of six simple one piece patterns repeated. The experimental program E-prime was used to design the study and display and capture the data. The program randomly assigned a dot matrix to every MP inference in the experiment. Consequently, each participant had a different randomly assigned set of matrices paired with each of the conditional arguments presented.

Procedure

Participants were tested in groups of four and were randomly assigned to the high or low load conditions. They were tested in partitioned booths and presented with instructions on a computer screen. Participants were informed that they would have to complete 64 reasoning problems and were instructed to answer either according to their beliefs or according to logic. Belief instructions emphasised the requirement to answer in relation to their knowledge of what is true in the world and indicate whether the conclusion was believable or unbelievable. They were then presented with following example:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be full?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world, you know that if you finish your drink then the glass will be empty, therefore the conclusion is unbelievable.

Logic instructions asked participants to assume each statement was true (even if in reality it was not true) and indicate whether the conclusion followed validly from preceding sentences. They were then presented with same example under logic instruction:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be empty?

s) VALID k) INVALID

In this particular example the correct answer according to logic is INVALID, because the first premise states that "if you finish your drink then the glass will be full" and supposing you "finish your drink" you must logically conclude that your glass will be full. This is why the conclusion "does it follow that the glass will be empty" is logically invalid.

The next set of instructions informed participant's that they would be presented with a dot pattern (for a short period) before each reasoning problem, which they had to remember and reproduce on an empty matrix grid presented to them after completing the reasoning question. They had eight practice trials to complete before the main experiment, four without a dot pattern and four with. After each practice trial, feedback was provided specifying whether they had answered correctly or not on the reasoning task. Depending on their mistake the feedback resembled the following:

You responded "invalid" instead of "valid". You were supposed to respond according to LOGIC. If you do not understand your mistake then ask the experimenter for further instructions before the end of the practice trials.

When participants started the experimental trials, the accuracy and latency data was captured for each trial and logged from the presentation of the full problem until a response was given.

2.2.2 Results

Following Handley et al. (2011), we analysed the full data and a reduced data set by eliminating those participants that scored below chance on the conflict problems (i.e. less than 50% accurate), which suggested that they were applying an inappropriate instructional set to the questions. We focus on the reduced data throughout this thesis (see Appendix C for full data sets) and dropped six participants from the current experiment giving a total of N = 75. All the accuracy data was Arcsine transformed to improve homogeneity of variance and control for the impact of ceiling effects (Milligan, 1987). Analyses on the latency data were performed on correct responses only and any missing data was replaced with the overall cell mean for that question type (i.e. belief conflict, belief no-conflict). Items with latencies greater than 2 Standard Deviations (SD) from the mean were removed in order to reduce the influence of outliers that skew the distribution. The tables in each chapter present latencies with outliers removed and percentage accuracy scores, prior to transformation.

A 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(High/Low load) mixed design was used with *instruction* (Belief/Logic) and *problem* type (Conflict/No-conflict) as within-subject factors and load as a between-subject factor. A repeated measures ANOVA (analysis of variance) on transformed accuracy produced a marginal main effect of *instruction*; $F(1, 73) = 3.435$, $p = .068$, $\boldsymbol{\eta_p^2} = .045$, indicating that performance was slightly better under logic instruction than belief instruction (*94% vs. 93%*). There was a main effect of *problem* type, $F(1, 73) = 51.057$, $p < .001$, $\boldsymbol{\eta_p^2} = .412$, showing accuracy scores significantly higher for no-conflict items compared to conflict items (*97% vs. 90%*). Interestingly there was *no* main effect of *load*, $F(2, 73) = 1.102$, $p = .297$, $\boldsymbol{\eta_p^2} = .015$, showing that overall performance was unaffected by the dot memory task.

There was a marginally significant interaction between *instruction* and *problem* type, $F(1, 73) = 3.717$, $p = .058$, $\eta_p^2 = .048$, identifying a significantly larger difference in performance for *problem* type under belief instruction (*89 % conflict vs. 97 % no-conflict*) than under logic instruction (*91% conflict vs. 97% no-conflict*). These results replicate Handley et al. (2011) where belief-logic conflict has shown a greater impact on the believability of the conclusion. However, *instruction* did not interact with *load*; $F(1, 73) = .020$, $p = .889$, $\eta_p^2 < .001$, indicating that high and low memory load did not have differential impact on belief and logic instructions. None of the remaining interactions were significant (all *p*'s > .10).

Table 2.3
Average Accuracy and Latency scores for Conflict and No-conflict Problems under Belief and Logic Instruction across both groups. Result exclude below chance scores and include correct only latencies. Experiment 1: N = 75

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
High Load Response Accuracy (%)	87	96	88	97	92
Latency Scores (ms)	4,199	3,585	3,621	3,327	3,683
Low Load Response Accuracy (%)	90	98	93	97	95
Latency Scores (ms)	3,948	3,655	3,727	3,449	3,695
Mean Accuracy (%) (across each cell)	89	97	91	97	
Mean Latency (ms) (across each cell)	4,074	3,620	3,674	3,388	

A repeated measures ANOVA was carried out on correct response latency data, which generated a main effect of *instruction*, $F(1, 73) = 36.921$, $p < .001$, $\eta_p^2 = .336$, with belief judgments taking more time to complete than logic judgments (*3,847ms vs. 3,531ms*). There was also a main effect of *problem* type; $F(1, 73) = 36.928$, $p < .001$, $\eta_p^2 = .336$, with conflict items taking longer than no-conflict items (*3,874ms vs. 3,504ms*) but again there was no main effect of *load*; $F(2, 73) = .005$, $p = .942$, $\eta_p^2 = .000$. There was, however, a marginal interaction between *instruction* and *load*; $F(2, 73) = 3.841$, $p = .054$, $\eta_p^2 = .050$, showing a bigger difference between latencies for belief and logic instruction in the high load condition (*3,892ms-Belief vs. 3,474ms-Logic*) compared to the low load condition (*3,802ms-Belief vs. 3,588ms-Logic*). There were no other significant interactions to report (all *p*'s > .10).

2.2.3 Discussion

Generally, there was a very high accuracy rate amongst the participants in Experiment 1. Nevertheless, the trend and direction of the accuracy data produced similar results to those of Handley et al. (2011) demonstrating that belief-conflict items produce less accurate responses and take longer to complete than logic-conflict items. The fact that conflict items take longer to complete implies that the participants, on some level, were aware of the conflict between logic (the form of the argument) and their beliefs regarding the content of the argument. This is consistent with previous research suggesting that participants are aware of conflict (e.g. De Neys & Glumicic, 2008).

Interestingly, the secondary load had no impact on reasoning accuracy in contrast to previous research carried out on syllogisms (De Neys, 2006). One explanation for the absence of load impact could be down to the simplicity of the secondary task and its insufficient demands on WM; however the results do suggest that the secondary task was adequately demanding with an average of 53% correct in the high load condition (complex matrices). An alternative explanation is that participants traded performance on the secondary task for performance on the primary task. However, 53% correct on dot matrices is substantially above the level of chance (0.79% for the complex matrices and 33.33% for the simple matrices) suggesting that participants were sufficiently engaged in the secondary task.

An important question at this point is why the dot memory task did not impact on the processes required to answer the reasoning questions? Perhaps we need to consider the possibility that the reasoning task and the dot memory task are drawing on distinct processing resources. According to authors such as Rabbitt (1997), executive function is comprised of a fragmented system of control processes; therefore the control of verbal and visuo-spatial processing might activate distinct executive resources. Consequently, it is quite possible that the specific type of secondary task used here, was in fact tapping into a different type of working memory resource to that required to evaluate the believability of presented conclusions. Capon, Handley & Dennis (2003) showed that syllogistic reasoning loaded both spatial and verbal working memory (VWM) whilst Handley, Capon, Copp & Harper (2002) showed that conditional reasoning only correlated with VWM, suggesting that conditional inferences may be more heavily dependent on VWM than syllogisms. Consequently, if the dot memory task demands more on spatial WM, this could explain the absence of load impact in Experiment 1. Shah and Miyake (1996) also suggested a separation of WM resources for spatial thinking and language processing. They interpreted their results as demonstrating how cognitive resources that process and sustain spatial and verbal information are likely to be distinguishable (Logie, 1995). For example, data from dual-task models have demonstrated how the maintenance of spatial information in WM can be disrupted by a simultaneous spatial task but not by visual tasks and vice versa (Logie, 1986; Logie & Marchetti, 1991). Kim, Kim and Chun (2005) further demonstrated how the type of interference shown on a Stroop test, depended on the type of WM load they used; in their words, impairment of executive control depends on the information of the concurrent task overlapping with the content of the WM.

There is neurophysiological evidence that different types of information (spatial and visual in this example) are processed through distinct neural pathways in the brains of primates (Mishkin, Ungerleider & Macko 1983). Subsequent research also provides neurological evidence that the visuo-spatial sketch pad can be further divided into two functionally and anatomically distinct systems (Bloom, 1956) and brain-imaging scans demonstrate that retention of spatial information in the short-term is managed at a different location to the visual information in the brain (Smith et al., 1995). Furthermore, Logie (1995) proposed a rather passive role for the spatial subsystem. Later research however, infers a fainter distinction between the WM and Storage-orientated Short Term Memory (STM) tasks (such as a dot memory task) and implicates the use of executive functions in both of them (Miyake, Friedman, Rettinger, Shah & Hegarty, 2001). Miyake et al. (2001) did, however, suggest that there are 3 spatial ability factors and they differ in the degree of executive involvement. It could be that the visuo-spatial storage is used when completing the dot memory task but does not place a substantial demand on the executive component. Consequently, a secondary task that burdens more heavily on executive functions rather than spatial STM may be a more appropriate secondary load. In the same paper Miyake et al. (2001) demonstrated a distinct separation between the Dot Memory task and Random number Generation (RNG). Their research indicated that whilst the two tasks correlated (inter-correlation: 08 = very small) they were in fact tapping into different functions. Essentially, the dot memory task taps into the Visuo-spatial STM-WM and RNG engages executive functioning. This became the rationale for using RNG as the secondary load in Experiment 2.

2.3 Experiment 2

Research has provided evidence that different types of information require the functioning of distinct executive resources (Logie, 1995; Shah & Miyake, 1996). This highlights the need to ensure the secondary task used draws upon the same cognitive resources as the primary task. There is the possibility that the dot memory task used in De Neys' (2006) research required the same executive resources as those involved in syllogistic reasoning but not the same as those necessary to complete conditional inferences. For example, Ford (1995) and Bacon, Handley and Newstead (2003) used written and verbal protocols to show that people use two distinct strategies for reasoning with syllogisms, some reason primarily spatially and others reason verbally. Spatial reasoners often use diagrams similar to Euler circles whilst verbal reasoners tend to approach syllogisms as if solving an equation, the principle of which was likened to the logical rules used to solve Modus Ponens and Modus Tollens arguments. Perhaps one explanation for De Neys' findings was that some of the participants used were primarily spatial reasoners, hence the significant effect the spatial load had on the syllogistic task. Whereas there is evidence to suggest that for conditional inferences, reasoning requires an abstract Working Memory (WM) medium for representation and does not require the Visuo-Spatial Sketch Pad (Handley et al., 2002; Toms, Morris & Ward, 1993). Therefore, a secondary task that places a substantial demand on the executive component of WM such as Random Number Generation (RNG) may have a greater impact upon the primary task.

Random Number Generation has been one of the most extensively used tasks to explore the functioning of the central executive component of the WM (Miyake et al., 2000). Typically the task requires participants to generate a sequence of random numbers from a set between 1 and 9, every second. Their performance is then measured through a variety of indices that calculate the randomness of the sequence generated. The ability to create a random series of numbers is said to demand heavily on WM, take for instance the conscious restraint needed to prevent the instinctive inclination to follow a number one with a number two, whilst maintaining an awareness of previously generated numbers to compare with the concept of what is random. Furthermore, it has been well established that the generation of random responses is disrupted by and impedes simultaneously performed, attention-demanding, secondary tasks (Baddeley, 1986; Brown, Soliveri & Jahanshahi, 1998).

In order to eliminate the possibility of number articulation being accountable for any interference in Experiment 2, the Articulatory suppression (AS) technique was used as the low load condition. Articulatory suppression requires participants to repeatedly say a word or number out aloud in order to load the phonological loop component of WM, by preventing silent verbal rehearsal (Baddeley, 1986). It has been widely used to examine the role of verbal rehearsal in cognitive tasks, and is recognised to interfere with verbal short-term memory (Baddeley, Lewis & Vallar, 1984), without interfering with the accuracy of conditional reasoning problems (Evans & Brooks, 1981).

Using the same methodology for the reasoning task as Experiment 1 and a secondary load more demanding of executive resources, we theorise that judgments requiring Type 2 processing will be disrupted by the engagement in random number generation. In accordance with the PC model, the secondary load should interfere more with performance on belief based judgments. Alternatively, the DI account would predict that RNG will have its greatest impact on logic based inferences.

2.3.1 Method

Participants

A total of 74 psychology undergraduates and post graduates from Plymouth University took part in Experiment 2. Fifty four female and 20 male participants were recruited and awarded course credits for participation, or gratitude for volunteering.

Design & Material

A 3 x 2 x 2 mixed design was used where participants were randomly assigned to Random Number Generation group (high load condition), Articulatory Suppression group (low load condition) or the Control group (no load condition). Each participant received the same 64 reasoning problems used in Experiment 1, under both belief and logic instructions (Instruction type). They consisted of an equal number of conflict and no-conflict problems and four practice trials (see experiment 1 for details).

Random Number Generation

The task of random number generation was based on the methodology used by Miyake et al. (2000) (although analysing random generation dates back to Baddeley, 1966). Participants were instructed to say aloud a number between 1 and 9, every second for the entire duration of the experiment. They were instructed to continue generating random numbers whilst solving the reasoning problems. To ensure they understood the concept of `random', they were given the following example:

Suppose you had written the numbers 1-9 on pieces of paper and put them in a hat. You take out one piece of paper, call out the number and return it to the hat. Then you reach for another piece of paper and do the same thing. The series of numbers you call out in that way should be random (Horne, Evan & Orne, 1982).

Before the start of the experiment, participants were told to keep the generation of numbers continuous, albeit tempting to pause whilst reading the questions they must keep to generating a number every second to the sound of the metronome beat. They were also advised that if they accidently went beyond the number range 1-9, they should try *not* to pause and continue on.

Articulatory Suppression

In the AS condition participants were instructed to say aloud the number two, every second for the duration of the experiment. The importance of keeping to the second was highlighted in the instructions.

Procedure

Participants were tested in maximum groups of four and were randomly assigned to the three conditions. They were tested in partitioned booths behind closed doors to keep vocalised distractions to a minimum. Each participant wore closed cup earphones to reduce background noise and listen to a metronome beat that was set to click every second. They each had a dicta-phone placed in front of the keyboard to capture the articulation of numbers. The control condition only had the reasoning task instructions to adhere to, whilst the RNG and AS condition had separate instructions for the secondary task (see Appendix A for details). Response latencies were logged from the presentation of the full problem until a response was given.

Analysing Randomness

The *RGCalc* program (Towse & Neil, 1998) was used to analyse the randomness of the numbers generated by each participant in the high load condition. The program produces many different indices needed to examine randomness, but the three main measures used here were taken from Towse & Valentine (1997) and are as follows:

R = Redundancy: A score of 0% means each response alternative is given with equal frequency where as 100% means the same response has been constantly selected throughout, for example: 2, 2, 2, 2.

RNG = Random Number Generation: measures how often a response alternative follows another (on a scale of 0 to 1) for example; how often 1 follows 7 or 4 follows 8, throughout the data set. The closer RNG is to 1 the more predictable the pair sequence.

A = Adjacency: measures a specific type of sequential pairing, in contrast to RNG analysis, for example; the commonality of 1, 2 or 3, 2 (neighbouring pairs on the number line). A score of 0% means no neighbouring pairs were presented, whilst 100% means all successive responses were adjacent number values.

The three measures were used to ensure participants adequately engaged in the secondary task. A score of 50% or higher on 2 out of 3 of the random indices was set as the criterion for eliminating participants not sufficiently engaging with the RNG task. Only one participant was removed on the basis of their randomness scores, for producing the sequence: 1, 2, 3, 4, and 5, throughout the experiment. Another elimination factor was the speed at which participants generated the random numbers. Participants were instructed to keep to one number per second. We calculated their articulation rate by predicting the number of digits that would have been generated in the task assuming an articulation rate of 1 per second. We then divided this number by the actual number of digits generated throughout the experiment. The criterion for eliminating participants was to remove those that took longer than two seconds per number; however no participants were eliminated on these grounds.

2.3.2 Results

A total of seven participants were removed (N = 67), six for scoring below chance on the *conflict* problems and one for producing a deliberate sequence of numbers and poor randomness scores. The accuracy data was Arcsine transformed and outliers were removed from the latency data. A mixed design ANOVA was carried out on the 2(Belief/Logic) x 2(Conflict/No-Conflict) x 3(RNG/AS/Control) mixed design. Accuracy scores revealed no main effect of *instruction;* $F(1, 64) = .973$, $p = .328$, $\eta_p^2 = .015$, indicating no significant difference between belief and logic judgments *(85% vs. 87%)*, but there was a main effect of *problem* type; $F(1, 64) = 50.555$, $p < .001$, $\eta_p^2 = .441$, showing poorer performance on conflict items (*81% vs. 90%*). *Load* generated a significant main effect; $F(3, 64) = 17.097$, $p < .001$, $\eta_p^2 = .348$, and post hoc analysis using LSD correction revealed significant differences in mean accuracy scores between the RNG and AS condition ($p = .018$), between the RNG and Control condition ($p < .001$) and a significant difference between the AS and the Control condition ($p = .001$). The means suggest that RNG was producing the greatest reduction in accuracy scores (*meanRNG = 78%, meanAS = 85%, meanControl = 93%*).

There was no interaction between *instruction* and *problem* type; $F(1, 64) = .697$, $p = .407$, $\eta_p^2 = .011$, but there was a significant interaction between *instruction* and *load;* $F(2, 64) = 5.679$, $p = .005$, $\eta_p^2 = .151$ (see Table 2.4).

Table 2.4

Average Accuracy and Latency scores for Conflict and No-conflict Problems under Belief and Logic Instruction, across 3 Conditions. Result exclude below chance scores and include correct only latencies. Experiment 2: RNG (N = 21), AS (N=22), Control (N = 24)

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
RNG					
Response Accuracy (%)	73	85	74	81	78
Latency Scores (ms)	4,391	3,960	3,605	3,623	3,895
AS					
Response Accuracy (%)	78	89	82	90	85
Latency Scores (ms)	3,891	3,434	3,481	3,333	3,535
Control					
Response Accuracy (%)	87	96	92	98	93
Latency Scores (ms)	4,182	3,648	3,731	3,534	3,774
Mean Accuracy (%) (across each cell)	79	90	83	90	
Mean Latency (ms) (across each cell)	4,155	3,681	3,606	3,497	

In order to determine between which groups the *instruction* by *load* interaction occurred, we reviewed two groups at a time using a repeated measures ANOVA. First comparing RNG with AS, the results revealed a main effect of *load;* $F(2, 41) = 4.768$, $p = .036$, $\eta_p^2 = .102$; with the RNG group producing poorer accuracy scores than the AS group (*78% vs. 85%*). There was also a marginally significant interaction between *instruction* and *load*; $F(1, 41) = 2.932$, $p = .094$, $\eta_p^2 = .067$, revealing that performance was better for logic-based judgments in the AS condition compared to those generating random numbers. An independent sample t-test on *instruction* confirmed that the significant difference between the AS and RNG condition was for logic instruction; $t(41) = 2.234$, $p < .05$, *AS* (M = 86.1, SD = .123), *RNG* (M = 77.8, SD = .124) compared to no difference for belief instruction; $t(41) = 1.055$, $p = .298$, *AS* (M = 83.4, SD = .141), *RNG* (M = 79, SD = .14).

Comparing RNG to the Control group, the outcomes were comparable, showing a main effect of *load;* $F(2, 43) = 52.089$, $p < .001$, $\eta_p^2 = .548$, with performance accuracy significantly better in the control condition and a significant interaction between *instruction* and *load;* $F(1, 43) = 4.690$, $p = .036$, $\eta_p^2 = .098$, again indicating that RNG had a greater effect on logic *(78% RNG vs. 95% control)* than belief *(79% RNG vs. 92% control)*. Follow up analysis revealed that scores in the RNG group differed significantly from the control group, for both logic judgments; $t(44) = 6.604$, $p < .001$, *RNG* (M = 76.9, SD = .124), *Control* (M = 95.0, SD = .047) and belief judgments; $t(44) = 4.020$, $p < .01$, *RNG* (M = 78.9, SD = .134), *Control* (M = 91.7, SD = .073). Finally, comparison between AS and the Control group produced a main effect of *load*; $F(2, 44) = 10.218$, $p = .003$, $\eta_p^2 = .188$; establishing better accuracy scores when no load was applied, whilst the absence of *instruction* by *load* interaction confirmed that the effect of articulation on performance did not vary as a function of instruction. In other words, there was no significant difference between instruction type for the Control and AS group, implicating RNG as having the main impact on instruction. All remaining interactions were not significant (all p's > .10).

A repeated measures ANOVA on correct response latency data uncovered a main effect of *instruction*; $F(1, 64) = 11.779$, $p = .001$, $\eta_p^2 = .155$, with belief judgements taking longer than logic judgments (*3,918ms vs. 3,551ms*). There was a main effect of *problem* type; $F(1, 64) = 9.098$, $p = .004$, $\eta_p^2 = .124$, with conflict problems taking longer (*3,880ms vs. 3,589ms*) but curiously there was *no* main effect of *load*; $F(2, 64) = 1.148$, $p = .324$, $\eta_p^2 = .035$, suggesting that irrespective of the level of load imposed, there was no significant difference in response times. Finally, the only significant interaction was between *instruction* and *problem* type; $F(1, 64) = 4.286$, $p = .042$, $\eta_p^2 = .063$, revealing that the effect of conflict was greater on beliefs (*4,155ms conflict vs. 3,681ms no-conflict)* than on logic judgments (*3,606ms conflict vs. 3,497ms no-conflict*) (see Appendix C for full data sets).

2.3.3 Discussion

In Experiment 2 we examined the impact of RNG on reasoning performance under belief and logic instruction. A secondary load more demanding of the executive components of WM produced the main effect we expected on overall accuracy scores. The results revealed a significant difference between the three conditions for their accuracy scores under belief and logic instructions; more specifically that there was no difference in accuracy on belief judgments between RNG and AS conditions, but there was a difference on logic judgments. However, although the comparison between the Control and the RNG condition revealed an interaction between *load* and *instruction* there was a significant difference for both types of judgment. Taken together these findings suggest that although there is some impact of RNG on belief judgments it has its greatest effect on logical judgments.

Experiment 1 revealed no effect of the dot memory task on the reasoning data whilst Experiment 2 establishes RNG as having its greatest effect on the ability to reason under logic instruction more so than belief. This finding is consistent with the DI model (see Figure 2.1) that suggests logic judgments are more demanding of Type 2 processing and involve more cognitive effort than judgments requiring a belief based response, thus, logic judgments are affected to a greater extent by the taxing secondary load. This is surprising given that belief judgments take longer and are influenced to a similar (in Experiment 2) or greater degree (Experiment 1) by conflict. An important question is how we reconcile these two apparently contradictory findings?

One potential explanation for the results concerns the presentation of the reasoning problems, and how it could be making logic judgments harder to solve. Take the following conditional statement:

If a child is crying then it is happy (Major Premise)

Participants are first presented with the major premise, which disappears when they press the space bar to reveal the categorical premise and conclusion:

The child is crying (Categorical Premise)

Is the child sad? (Conclusion)

Reasoning on the basis of logic, requires the integration of information from the major premise with the categorical information of the second premise and then the evaluation of the conclusion. When participants are asked to evaluate the validity of a conclusion, they are expected to have retained and subsequently recall the information from the major premise, creating extra demands on WM; whereas reasoning on the basis of beliefs requires integration of information from the categorical premise with that presented in the conclusion which always remains on the screen. Consequently, the greater effect of RNG on logic judgments could be the result of participants forgetting the first premise which is only problematic when the secondary task taxes heavily on WM resources. In order to determine whether RNG impedes on logical reasoning by interfering with the WM process of integrating premise information or whether participant are failing to remember the information presented in the major premise, in

Experiment 3 we modify the procedure, such that the first premise always remains on the screen.

2.4 Experiment 3

In the previous experiment judgments based on the validity of the argument required retention of the information presented in the major premise, after it disappeared from the screen, and the subsequent recall of that information when answering the question. For belief based questions only the categorical premise and conclusion are required for judgment, which remained on the screen throughout response generation. Consequently, the outcomes from Experiment 2, implying that RNG had more of an impact on logical inferences, could be explained by the added WM load of remembering and recalling the major premise. To investigate this possibility the following experiment will keep the major premise on the screen in each trial which will limit the memory requirement under logical instruction in order to match the WM load in each condition.

The results from Experiment 2 showed that carrying out a secondary task such as articulatory suppression (AS) was more challenging than completing the primary task alone, as in the control condition. However, the accuracy scores between the two groups confirmed that articulation was not responsible for the effects observed and did not impact differentially on performance. Therefore we include RNG and the Control condition for comparison in Experiment 3 and remain interested in determining whether the secondary load will have its greatest impact on belief or logic judgments.

2.4.1 Method

Participants

A total of 76 paid participants took part in Experiment 3, 45 were female and 31 were male, all of whom were paid £4 each for their time.

Design & Materials

A 2 x 2 x 2 mixed design was used where participants were randomly assigned to the RNG or Control condition. Each participant received the same 64 reasoning problems as Experiment 1 and 2. Thirty two problems were presented under belief instruction and 32 under logic instructions (Instruction type), consisting of conclusions that were valid and unbelievable or invalid and believable (Conflict) or conclusions that were valid and believable or invalid and unbelievable (No-conflict) (see experiment 1 for details).

Reasoning Task:

In each trial the major premise was presented on the computer screen for a total of 3000ms and remained on the screen to ensure participants had enough time to read and understand it before the categorical premise, conclusion and response options were presented.

Random Number Generation

We used the same procedure as Experiment 2.

Procedure

The same instructions and materials as Experiment 2 were used, but this time participants were randomly assigned to one of two conditions: Random Number Generation or the Control condition.

Analysing Randomness

The *RGCalc* program (Towse & Neil, 1998) was used to analyse randomness (see experiment 2).

2.4.2 Results

We removed the data from 18 participants, three for taking longer than two seconds to generate numbers and 15 for scoring below chance on the conflict problems (i.e. less than 50% accurate). In total five were removed from the Control condition (N = 30) and 13 from the RNG condition (N = 28) giving a total of *N = 58*. All accuracy data was Arcsine transformed and outliers were removed from the latency data. Only correct responses were recorded and exclude the 3000ms presentation time of premise 1. Any missing data was replaced with the overall cell mean for latency scores (the full data sets are presented in Appendix C).

A 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(RNG/Control) mixed design ANOVA on the accuracy scores revealed a main effect of *instruction*; $F(1, 56) = 14.460$, $p < .001$, $\boldsymbol{\eta_p^2} = .205$, with logic judgments producing the most accurate scores across both conditions (*90% vs. 84%*). There was a main effect of *problem* type; $F(1, 56) = 54.650$, $p < .001$, $\eta_p^2 = .494$, showing poorer performance on conflict items (*81% vs. 94%*). There was also a marginal main effect of *condition;* $F(2, 56) = 3.874$, $p = .054$, $\boldsymbol{\eta_p^2} = .065$, which showed that RNG had an overall effect on performance compared to the control condition (*84% vs. 90%*). However, there was no interaction between *instruction* and *condition*; $F(1, 56) = .019$, $p = .891$, $\eta_p^2 < .001$, showing that RNG did not impact on logic judgments (*88% RNG vs. 92% control)* more so than belief judgments *(81% RNG vs. 87% control)*. The remaining interactions were not significant (all *p'*s > .10).

For the latency scores there was no main effect of *instruction;* $F(1, 56) = .000$, $p = .992$, $\boldsymbol{\eta_p^2} = .000$, but conflict problems took longer to complete than no-conflict problems (*7,360ms vs. 6,923ms*) producing a significant main effect of *problem* type; $F(1, 56) = 5.661$, $p = .021$, $\boldsymbol{\eta_p^2} = .092$. There was a main effect of *condition*; $F(2, 56) = 17.579$, p

< .001, η_p^2 = .239, demonstrating a significant increase in latency scores when participants were engaged in RNG (*8,375ms vs. 5,907ms*). There was an interaction between *instruction* and *problem* type, $F(1, 56) = 8.414$, $p = .005$, $\eta_p^2 = .131$, which follow up analysis revealed that the impact of conflict was significant for belief judgments; $F(1, 56) = 14.896$, $p < .001$, $\eta_p^2 = .210$, but not logic judgments; $F(1, 56) = .065$, $p = .799$, $\eta_p^2 = .001$. However, *instructions* did not interact with *condition*; $F(1, 56) = .920$, $p = .341$, $\eta_p^2 = .016$, therefore latencies for both belief and logic judgments were comparable across both conditions. Finally, none of the remaining interactions were significant (all *p*'s > .10).

Table 2.5

Average Accuracy and Latency scores for Belief and Logic Instructions across both Conditions. Result exclude below chance scores and include correct only latencies. Experiment 3: RNG (N = 28), Control (N = 30).

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
RNG					
Response Accuracy (%)	70	91	82	93	84
Latency Scores (ms)	8,813	7,643	8,219	8,826	8,375
Control					
Response Accuracy (%)	82	92	87	97	90
Latency Scores (ms)	6,505	5,509	5,903	5,711	5,907
Mean Accuracy (%) (across each cell)	76	92	85	95	
Mean Latency (ms) (across each cell)	7,659	6,576	7,061	7,269	

* RNG = Random Number Generation

2.4.3 Discussion

The results from Experiment 2 showed that RNG had its greatest impact on logic based judgments. In order to establish whether this finding was the result of increased WM demands affecting logic based reasoning, Experiment 3 kept the major premise on the screen whilst participants completed each trial. This alteration controlled for the possibility that forgetting the information presented in premise one hindered logical performance under RNG conditions.

The findings in Experiment 3 demonstrated that even though overall accuracy scores were poorer in the RNG group, the secondary load did not differentially impact instruction type, as indicated by the absence of an interaction between *instruction* and *condition*. Currently, this suggests that simple MP inferences under both belief and logic instruction require some form of effortful processing, thereby contesting a DI interpretation of the findings. The general discussion will consider the implication of these findings with respect to the PC account.

2.5 General Discussion

The main objective of this chapter was to investigate the effect of a secondary load on reasoning performance when participants were required to make judgments based on logic or beliefs. Our aim was to examine the extent to which belief based judgments require effortful processing and evaluate the basic principles of the Default Interventionist (DI) account (see Figure 2.1), which claims that beliefs are triggered by default (Evans, 2006). The effect of secondary load would help us identify whether, belief, logic or both require effortful, Type 2 processing.

Experiment 1 broadly replicated the findings presented by Handley, Newstead and Trippas (2011) and concluded that overall performance on belief judgments was less accurate and look longer than logic judgments and that belief-logic conflict had a greater impact on the believability of the conclusion. However, we failed to replicate De Neys (2006) work which demonstrated that a dot memory task increased the effect of belief-bias with syllogistic inferences. These conflicting results may be due to the visuo-spatial attributes of the memory task and the strategies used to reason with syllogistic inferences. Mental model theory assumes that syllogisms are mentally represented within spatially structured models (Friedman et al., 2008) and there is evidence to suggest that they are visuo-spatial in nature (Friedman et al., 2006). This could explain the effectiveness of a spatially demanding secondary task in De Neys' research. In addition, other research has demonstrated that some people apply spatial strategies when reasoning with syllogisms, whilst others employ verbal strategies (Bacon, Handley & Newstead 2003; Ford, 1995). If participants where predominately spatial reasoners in De Neys study then a dot memory task would have been adequately demanding on syllogistic performance.

In addition, Reverberi et al. (2010) suggest that syllogistic reasoning involves qualitatively different cognitive processes to conditional "if *p* then *q*" arguments. Through research on brain activation during the information integration stage associated with the second premise, they showed that the processes recruited when reasoning syllogistically were not engaged by conditional reasoning. Therefore, another explanation for the lack of effect of load could be that the dot memory task taps into distinct executive functions to those required for solving conditional inferences. In effect the task could have drawn heavily on visuo-spatial components, affecting syllogistic reasoning rather than the executive components required for conditional inferences such as verbal working memory (Handley et al., 2002). Consequently, the lack of overlap between the primary and secondary task could have resulted in limited taxation of WM resources resulting in the absence of an effect of load (Logie, 1986; Logie & Marchetti, 1991).

In both Experiments 2 and 3 we used a Random Number Generation task (RNG) for its well established capacity to engage the central executive element of WM and its ability to interfere with conditional reasoning (Meiser, Klauer & Naumer, 2001). In Experiment 2 we observed that differential memory demands of each judgment, as a consequence of the first premise disappearing, could have explained the selective impact of RNG on logic judgments. Consequently, in Experiment 3 we controlled for this by keeping the major premise on the screen throughout the reasoning process which produced an overall effect of load on accuracy. These findings demonstrate that RNG impacts equally on both types of judgments, suggesting they both require some form of effortful processing. Furthermore, the results are inconsistent with the DI account of reasoning that suggests belief judgments are fast, Type 1 processes, given that they are affected by cognitive load and take longer to process than logic judgments.

One of the assumptions made at the beginning of this chapter, based on Handley et al.'s findings, was that both belief and logic based responses are triggered simultaneously but logic completes first (see Figure 2.2). The logical response then requires inhibiting in order for a belief based output to complete. Consequently, if RNG impacts on the cognitive resources required to inhibit logical responses then load would have had its greatest impact on belief judgments. The results from Experiment 3 revealed an equal impact of load, suggesting Type 2 processing under both instructional sets. However belief judgments were more prone to error and generally took longer to complete which may suggest distinct executive resources required for belief and logic based reasoning. According to the PC model, belief judgments require inhibition resources to supress fast completing responses based on logical validity, but it is also possible that logic judgments require some reflective Type 2 processing such as premise integration, model construction and conclusion generation.

This chapter used simple Modus Ponens (MP) arguments because they allow us to evaluate the extent to which belief-logic conflict impacts on logical or belief based judgments. However, it has been argued that MP is an intuitively direct deductive inference that can be triggered by subliminal stimuli and carried out automatically, even without instruction (Reverberi, Burigo & Cherubini, 2009; Reverberi, Pischedda, Burigo & Cherubini, 2012; Rips, 1988; 1994). The current findings do not support MP implicit nature since RNG impacted on overall performance in Experiment 3. However, the effect of secondary load was marginal therefore it is important to extend these findings to more complex arguments that are claimed not to be drawn automatically (Reverberi et al., 2012).

Another possible explanation for the marginal effect of load and the higher logic scores produced in Experiment 3 is that participants were not engaging in formal deductive reasoning, but simply developing a shortcut strategy that allowed them to bypass the requirements to engage in more effortful processing; one such strategy is known as Matching Bias. In MP arguments a valid conclusion will be a direct match to the content of the consequent in the major premise. This potentially provides a shortcut for making a validity judgment that would not require integration of the minor premise. In other words, participants could have provided an accurate `valid' response when the content of the first premise matched the content of the conclusion and an accurate `invalid' response for mismatched content. Such a shortcut could explain why participants responded more rapidly under logic instructions and were only marginally affected by the secondary task. We discuss the suitability of disjunctive arguments for eliminating this matching shortcut, in the next chapter.

In Chapter 3 we investigate whether our findings can extend to more complex disjunctive inferences. Handley et al. (2011) replicated their initial findings using disjunctive arguments; additionally Reverberi et al. (2012) found that disjunctive arguments were not automatic, implying that they require more effortful processing than MP. Using more complex disjunctive arguments we expect to find that they are harder to reason with (Johnson-Laird, Byrne & Schaeken, 1992) and will produce lower accuracy scores over all. With regards to instruction type, we can expect that once again both belief and logic will be affected by the generation of random numbers. Finally, another key issue to be addressed in the following chapter is the extent to which belief/logic conflict impacts on judgments as a function of the complexity of the argument.

Chapter 3: The effect of RNG on simple and complex judgments.

3.1 Introduction

At the end of Chapter 2 the conclusion drawn was that simple Modus Ponens (MP) arguments for both belief and logic based judgments required *some* effortful processing. However, we also discussed two possible explanations for enhanced logic based accuracy scores that require further investigation. The first was Matching bias; a phenomenon documented in the selection task (Wason, 1968), connected with the problem of understanding implicit negation (Evans, 1972). In the selection task, participants are presented with four cards, each displaying one of two letters (A & D) or one of two number (3 & 7). The conditional rule presented to the participants is *"if there is an A on one side of the card, then there is a 3 on the other side of the card"* and they are required to judge which card they need to turn over in order to determine truth or falsity of the rule. The correct response is A and 7, since the statement can only be falsified by revealing a case where an A card does not have a 3 on the other side. However, in the majority of cases, people ignore negation and simply match the cards referred to in the propositional rule (A & 7) (Evans, 1998). One possibility is that logical judgments can be made using a shortcut based upon this matching strategy, for example the following:

If a child is crying then it is sad

Suppose the child is crying

Does it follow that the child is sad

With MP arguments a valid conclusion directly matches the content of the consequent in the major premise, offering a possible shortcut with logic based judgments and avoiding the requirement to integrate the minor premise. Furthermore, this matching strategy is more probable when the major premise remains on the screen as in Experiment 3. Disjunctions, on the other hand, do not permit this shortcut strategy, for example:

Either the sky is blue or it is pink

Suppose the sky is not pink

Does it follow that the sky is blue

Therefore, the first experiment in the present chapter will attempt to extend the previous findings by including complex disjunctives arguments.

A second possible explanation for our findings is that logical MP inferences are drawn automatically (Reverberi, Burigo & Cherubini, 2009; Reverberi, Pischedda, Burigo & Cherubini, 2012; Rips, 1988; 1994) and the possibility that they can be drawn automatically or by default could explain the greater effect of conflict on belief judgments. Reverberi et al. (2012) found that *disjunctive* arguments were not automatic and did not produce the same priming effects as MP inferences. Consequently, one might expect logic to interfere with beliefs for more automatic MP inferences but not with more complex logical arguments. The first study was also motivated by whether Random Number Generation (RNG) would impact on problems of increased complexity.

In the previous chapter we discuss two Dual Process models of reasoning; the Default Interventionist (DI) model (see Figure 2.1, Chapter 2) and the Parallel Competitive (PC) model (Figure 2.2, Chapter 2). According to the DI model, logic based judgments are more demanding of Type 2 processes than belief based judgments and should be more affected by a secondary load. In order to explain the greater accuracy scores on logic judgments in Experiment 3 we could conjecture that MP arguments are drawn automatically (Reverberi et al., 2012) and that RNG interferes with response generation as opposed to the effortful process of activating a relevant inference rule. This would also explain why RNG impacts on belief judgments. Therefore, the DI account would predict that by increasing logical complexity, RNG would have its greatest impact on logic judgments thus encouraging belief based responses and producing the traditional belief-bias effect. The PC model would also predict that logic would be more impacted on by RNG for complex arguments but also that the bigger impact of conflict on belief judgments, as displayed in Experiment 3, would be reduced.

3.2 Experiment 4

In the following experiment we continue with the use of RNG as our secondary load in order to replicate the findings from Experiment 3 but also examine whether RNG impedes on performance to a greater extent when complex logical arguments are introduced. Moreover, we present some individual differences (ID) measures for additional investigation.

Research on reasoning ability and ID has demonstrated that limitations in WM is key in explaining reasoning performance, with those higher in WM capacity, more proficient at drawing logical inferences (Capon, Handley & Dennis, 2003) and those with a higher WM span better equipped to use their resources to inhibit the activation of counterexamples when they conflict with the logical validity of a problem (De Neys et al., 2005). The degree of reasoning bias in various tasks has consistently correlated with individual differences, specifically cognitive ability (or general intelligence) as a result of its strong association with WM capacity (Conway, Kane, & Engle, 2003). Individuals with lower cognitive ability are usually more biased as a result of their limited WM capacity (Sá, West & Stanovich, 1999).

Cognitive style refers to an individual's motivation to engage in analytical processing (Pennycook, Cheyne, Seli, Koehler & Fugelsang, 2012) and thus predict that those with a lower propensity to reason analytically will rely on less effortful, heuristic processing. Cognitive style can be measured using self-report questionnaires such as the Actively Open-minded Thinking (AOT) scale (Macpherson & Stanovich, 2007; Stanovich & West, 1997; 2007) and the Cognitive Reflection Task (Frederick, 2005). AOT is a fundamental component of critical thinking, highlighting the ability to think in ways that support our own views as well as objectively search for alternative explanations in order to make impartial judgments. Thinking dispositions have been shown to correlate with an individual's ability to override belief-bias in syllogistic reasoning (Macpherson & Stanovich, 2007; Kokis, Macpherson, Toplak, West & Stanovich, 2002; Sá et al., 1999; Stanovich & West, 1998a). The CRT measures an individual's ability to suppress and disregard initial intuitive responses, in order to arrive at the correct, deliberated answer. The 3 item questionnaire is said to be a good predictor of the tendency to use heuristics and biases in making judgments and can explain more variance than the typical measures of cognitive ability, executive functioning and thinking dispositions (Toplak, West & Stanovich, 2011). For example the first of the items on the questionnaire reads;

"A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost?"

The intuitive response is often 10p; however, this would mean that the bat only costs 90p more than the ball. With further deliberation you can arrive at the correct answer which is 5p, making the bat, £1.05 and £1 more expensive than the ball. Toplak et al. (2011) offers the CRT as a measure that can assess the inclination towards `miserly processing', or the inclination to default to Type 1 processing due to its low computational expense (Toplak, West & Stanovich, 2014).

Experiment 4 will examine the relationship between cognitive style and ability on judgments made on the basis of logical validity and conclusion believability. We conjecture that if both belief and logic based judgments require some effortful processing then they should both correlate with WM measures. Furthermore, a higher ability to disregard an inappropriate intuitive response on the CRT should be linked with the instruction type that requires more effortful processing.

3.2.1 Method

Participants

A total of 80 participants, 65 females and 15 males, took part in Experiment 4 in exchange for course credits.

Design & Materials

A 2 x 2 x 2 x 2 mixed design was used where participants were randomly assigned to the RNG (Experimental) or Control condition. Each participant received 128 reasoning problems, both MP and disjunctive arguments (Complexity) under both belief and logic instructions (Instruction type), which consisted of both conflict and no-conflict problems (Problem type) (see Table 3.1 for more details).

Reasoning Task:

The same 64 MP conditionals from Experiment 3 were used in Experiment 4, with an additional 64 disjunctive arguments. The disjunctives consisted of an equal number of *denial inferences* (see Disjunctive Type A, Table 3.1) where the categorical premise denies one of the components of the major premise and an equal number of *affirmation inferences* which involves confirmation of one of the components to the denial of the other (see Disjunctive Type B). The content of the arguments had an equal number of problems under belief and logic instruction and an equal number of believable or unbelievable, valid or invalid arguments. Thirty two of the disjunctives were presented under belief instruction and 32 under logic instruction. In half the arguments both belief and logic corresponded, in the other half they were in conflict with each other (see Appendix B for materials).

Table 3.1 gives examples of conflict and no-conflict arguments associated with disjunctive Type A and B. The use of both denial and affirmation inferences ensures that there are no confounds between the polarity of the conclusion (i.e. whether it is negative or affirmative) and logical validity. The number of practice trials was also increased to include disjunctives arguments.

Table 3.1
Examples of the Disjunctive arguments used in Experiments 4.

Conflict Arguments	**No-conflict Arguments**
Disjunctive Type A	Disjunctive Type A
Argument 1 Either the sun is yellow or it is blue Suppose the sun is not yellow Does it follow that the sun is blue? Logic: ✓ Belief: X	*Argument 2* Either the sea is blue or it is pink Suppose the sea is not pink Does it follow that the sea is blue? Logic: ✓ Belief: ✓
Argument 3 Either the sun is yellow or it is blue Suppose the sun is not yellow Does it follow that the sun is not blue? Logic: X Belief: ✓	*Argument 4* Either the sea is blue or it is pink Suppose the sea is not pink Does it follow that the sea is not blue? Logic: X Belief: X
Disjunctive Type B	Disjunctive Type B
Argument 1 Either the sea is blue or it is pink Suppose the sea is pink Does it follow that the sea is not blue? Logic: ✓ Belief: X	*Argument 2* Either the sun is yellow or it is blue Suppose the sun is yellow Does it follow that the sun is not blue? Logic: ✓ Belief: ✓
Argument 3 Either the sea is blue or it is pink Suppose the sea is pink Does it follow that the sea is blue? Logic: X Belief: ✓	*Argument 4* Either the sun is yellow or it is blue Suppose the sun is yellow Does it follow that the sun is blue? Logic: X Belief: X

Random Number Generation

See Chapter 2, Experiments 2 and 3 for more details.

Measures of Individual Differences

Although cognitive ability and cognitive style are moderately to highly correlated, and both predict aspects of independent variance on heuristic and biases tasks (Stanovich, 1999), we measured both using the AH4 Group Test of General Intelligence: Part one (Heim, 1970) as a measure of ability and the AOT (Macpherson & Stanovich, 2007) and CRT (Frederick, 2005) as measures of style.

Part one of the **AH4** was the only section of the test administered since this is designed to assess the individual's verbal and numerical ability which correlates more with verbal working memory than the visuo-spatial measures of part two. Since both sections are highly correlated (Newstead, Handley, Harley, Wright & Farrelly, 2004) and verbal working memory has been shown to correlate with conditional reasoning (Handley, Capon, Copp & Harper, 2002) we decided that assessing the first section would suffice as the relevant measure of intelligence.

The **Active Open-Minded Thinking Scale** was taken from Macpherson & Stanovich, (2007) which comprises of 41 items, all taken from an assortment of sources: 10 items from a flexible thinking scale developed by Stanovich and West (1997); nine items measuring dogmatism (Paulhus & Reid, 1991); eight items from the Openness-Values facet of the Revised NEO Personality Inventory (Costa & McCrae, 1992); nine items from the belief identification scale developed by Sá et al. (1999); three items from the categorical thinking subscale of Epstein and Meier's (1989) constructive thinking inventory and two items from a counterfactual thinking scale developed by Stanovich and West (1997). The questionnaire required participants to respond using a six-point Likert rating scale.

The **Cognitive Reflection Task (CRT)** consisted of the following three questions:

- A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost?..............................
- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?............................
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?.................................

The correct answer for the bat and ball problem, is 5 pence, because the bat costs £1 more than the ball (i.e., the bat costs £1.05 and the ball costs 5p). In the widget problem, the correct response is five minutes and for the lily pads problem the correct answer is 47. Those who fail to engage in the small amount of effortful processing required to arrive at the right response, often give one or all of the following responses; 10p, 100 minutes and 24 days.

Analysing Individual differences:

Participants who scored at or above the median AH4 score of 42 were assigned to the high cognitive ability group (n =37), the remaining participants were assigned to the low ability group (n = 35). The same method was applied to the AOT scale, with those scoring 176 or above as the highly open-minded group (n = 38) and those below in the low open-minded group (n = 34). Finally, for the CRT task those that scored one or more out of three were assigned to the High CRT group (n = 29) and those that failed to give a correct answer were assigned to the Low CRT group (n = 43) (see Appendix B for ID materials).

Participants

Participants were tested in maximum groups of four and were randomly assigned to either the Control condition or Random Number Generation condition. Each participant was administered the AH4 test and given 10 minutes to complete the task. They then proceeded with the computer based reasoning task using the same instructions as Experiment 3 (Chapter 2) with additional practice trials containing disjunctive arguments (see Table 3.1). On completion of the reasoning task, they were asked to complete the AOT scale followed by the CRT, both of which were presented in paper format under no time restraints.

3.2.2 Results

Consistent with previous experiments, we eliminated those that scored 50% or below on the conflict items for both the RNG and Control condition. We also removed participants that took longer than two seconds on average to generate random numbers on the basis that they were not adequately engaged in the secondary task. In total, 13 participants were eliminated, three for exceeding an average of two seconds per number generation and two for producing stereotyped sequences which gave a total of N = 31 for the RNG condition and N = 36 for the Control condition.

A 2(Belief/Logic) x 2 (Conflict/No-Conflict) x 2 (MP/Disjunctives) x 2 (RNG/Control) mixed design ANOVA was carried out on accuracy measures which were Arcsine transformed. The analysis showed a main effect of *instruction* (Belief/Logic); $F(1, 65) = 19.266$, $p < .001$, $\boldsymbol{\eta_p^2} = .229$, with logic judgments producing the most accurate scores across both groups *(87% vs. 81%)*. There was a main effect of *problem* type

(Conflict/No-Conflict); $F(1, 65) = 66.225$, $p < .001$, $\eta_p^2 = .505$, with poorer performance on conflict items compared to no-conflict items *(78% vs. 90%)*. There was also a main effect of *complexity* (Disjunctives/MP); $F(1, 65) = 43.964$, $p < .001$, $\eta_p^2 = .403$, where MP produced higher accuracy scores than disjunctive judgments *(87% vs. 81%)* and a main effect of *condition*; $F(2, 65) = 13.193$, $p < .001$, $\eta_p^2 = .169$, demonstrating that accuracy was lower under load *(80% vs. 88%)*. However, load did not significantly interact with instruction; $F(1, 65) = 1.968$, $p = .165$, $\eta_p^2 = .029$.

There was also an interaction between *instruction* and *problem* type; $F(1, 65) = 20.278$, $p < .001$, $\eta_p^2 = .239$, showing that there was a bigger impact of conflict on belief problems (*72 % conflict vs. 90 % no-conflict*) than logic problems (*84% conflict vs. 90% no-conflict*). Follow up analysis was carried out to establish if the effects held for both arguments types and results revealed that for MP, *instruction* and *problem* type significantly interacted; $F(1, 65) = 9.613$, $p = .003$, $\eta_p^2 = .129$, with a bigger impact of conflict under belief instruction, which was also confirmed for disjunctive judgments; $F(1, 65) = 22.868$, $p < .001$, $\eta_p^2 = .260$. Mean accuracy scores can be found in Table 3.2. Finally there was a significant interaction between *complexity* and *problem* type; $F(1, 65) = 10.962$, $p = .002$, $\eta_p^2 = .144$, showing that the difference between conflict and no-conflict problems is larger for MP (*80% conflict vs. 93% no-conflict)* than disjunctives (*76 % conflict vs. 86 % no-conflict)*. No other interactions were significant (all *p'*s > .05).

Table 3.2

Average Accuracy and Latency scores across both groups for Belief and Logic Instructions, in each Problem Type, for MP and Disjunctive arguments. Experiment 4: RNG (N = 31), Control (N = 36).

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
RNG - MP					
Average Response (%)	68	91	81	91	83
Latency Scores (ms)	7,747	7,517	7,898	7,376	7,635
Control - MP					
Average Response (%)	79	94	90	97	90
Latency Scores (ms)	5,327	5,515	6,398	5,295	5,634
RNG - Disjunctives					
Average Response (%)	64	84	77	79	76
Latency Scores (ms)	9,443	9,518	9,652	8,945	9,390
Control - Disjunctives					
Average Response (%)	75	89	88	92	86
Latency Scores (ms)	7,431	7,552	6,878	6,781	7,161
Mean Accuracy (%) (across each cell)	72	90	84	90	
Mean Latency (ms) (across each cell)	7,487	7,526	7,707	7,099	

As in chapter 2, outliers were removed from the latency scores and missing data accounted for no more than 4% of the overall data in this chapter. A mixed design ANOVA was carried out, which produced no main effect of *instruction* (Belief/Logic); $F(1, 65) = .808$, $p = .372$, $\eta_p^2 = .012$, and no main effect of *problem* type (Conflict/No-Conflict); $F(1, 65) = 2.611$, $p = .111$, $\eta_p^2 = .039$. There was however, a significant main effect of *complexity* (Disjunctives/MP); $F(1, 65) = 66.329$, $p < .001$, $\eta_p^2 = .505$, and a main effect of condition; $F(2, 65) = 21.506$, $p < .001$, $\eta_p^2 = .049$, confirming that disjunctive judgments take longer to solve than MP *(8,276ms vs. 6,635ms)* and that participants take significantly longer to produce responses when engaged in a secondary task *(8,513ms-RNG vs. 6,398ms-Control)*. The analysis did not produce any significant interactions for the latency scores (all *p'*s > .05).

Accuracy scores on the ID measures were collapsed across both the experimental and control conditions. We only report significant main effects and interactions between experimental variables and ID measures and only excluded those that scored 50% or below on conflict items of the reasoning task, since performance on the secondary task was not relevant for this analysis. Between subject analyses was carried out on high and low groups for each ID measure which produced the following results: a significant main effect of AH4; $F(2, 70) = 9.607$, $p = .003$, $\eta_p^2 = .121$, with the high AH4 group performing better *(88%)* than the low group *(79%)* and a significant interaction between *instruction* and AH4; $F(1, 70) = 4.004$, $p = .049$, $\eta_p^2 = .054$, indicating that there was a larger difference between high and low cognitive ability for logic judgments *(92% high vs. 82% low)* compared to belief judgments *(84% high vs. 77% low)* (see Table 3.3).

Similarly the results from the CRT measure produced a significant interaction between *instruction* and *CRT*; $F(2, 70) = 4.024$, $p = .049$, $\eta_p^2 = .054$, showing a bigger difference between the high and low groups for logic judgments *(90% high vs. 85% low)* compared to belief judgments *(80% high vs. 81% low)* but there was no main effect of CRT; $F(2, 70) = .409$, $p = .524$, $\eta_p^2 = .006$. There was also a marginally significant interaction between *problem* type and *CRT*; $F(2, 70) = 3.235$, $p=.076$, $\eta_p^2 = .044$, with less of an effect of conflict in the high group (*81% conflict vs. 90% no-conflict*) compared to the low group (*77% conflict vs. 90% no-conflict*) (see Table 3.4). The results suggest that high CRT scorers perform better on logic based judgments and separate analysis on the high and low groups revealed that the difference between conflict and no-conflict items was larger for the low CRT group. This could imply that those who score high on the CRT measure have a more accomplished ability to reason based on the logical structure of an argument. Finally, there was no main effect of open-mindedness; $F(2, 70) = 1.551$, $p = .217$, $\eta_p^2 = .022$, and no significant interactions with AOT for accuracy scores, nor did the latencies significantly interact with any ID measure (see Appendix C for full data sets).

Table 3.3

Average Accuracy scores for Belief and Logic Instructions on Conflict and No-conflict items across High and Low AH4. Results exclude below chance accuracy scores. High AH4 (N = 37), Low AH4 (N =35).

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
High AH4 Average Response (%)	75	92	90	94	88
Low AH4 Average Response (%)	69	85	78	85	79
Mean Accuracy (%) (across each cell)	72	89	84	90	

Table 3.4

Average Accuracy scores for Belief and Logic Instructions on Conflict and No-conflict items across High and Low CRT. Results exclude below chance accuracy scores. High CRT (N = 29), Low CRT (N =43).

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
High CRT Average Response (%)	71	89	90	90	85
Low CRT Average Response (%)	73	89	80	90	83
Mean Accuracy (%) (across each cell)	72	89	85	90	

3.2.3 Discussion

In the introduction to this chapter we offered two possible explanations for greater accuracy on logic judgments in Experiment 3; one being a *matching* shortcut facilitated by the structure of MP arguments when instructed to reason on the basis of logic. The second refers to the implicit nature of MP inferences (Reverberi, Burigo & Cherubini, 2009; Reverberi, Pischedda, Burigo & Cherubini, 2012; Rips, 1988; 1994) and the possibility that they can be drawn by default. Research indicates that disjunctive arguments do not complete automatically (Reverberi et al., 2012) nor is any matching heuristic possible with these complex inferences. Therefore, if either explanation is responsible for the greater impact of conflict on belief judgments then the conflict should not arise with disjunctive arguments. Furthermore, we wanted to examine whether RNG would have greater impact when logical complexity was increased.

The results revealed there was no evidence that increasing complexity reduced the influence of conflict on belief judgments. This is inconsistent with an explanation of this effect based upon a matching strategy or the automaticity of MP inferences. The overall accuracy scores for logic judgments were better than belief judgments and the validity of the argument still had greater impact on the believability of a conclusion even for more complex inferences. Furthermore, the findings showed a robust effect of the secondary load for both disjunctive and conditional arguments under both sets of instructions. Again this suggests that belief based judgments are not available by default, contrary to the DI model, and both belief and logic inferences require effortful processing that pull on executive resources.

However, higher scores on both the CRT and AH4 tasks were associated with more accurate performance under logical instructions, whilst there was a smaller difference for belief based judgments. It is important to note that these effects were small, but nevertheless they suggest that cognitive ability and cognitive style are most strongly related to logical reasoning. One possibility is that distinct executive resources are required under different instructional conditions. In the general discussion we will consider the implications of these findings in more detail.

A characteristic of the experimental task is that participants are often required to switch from trial to trial between responses based upon logic or beliefs. There is good experimental evidence that task switching is demanding of executive resources (Monsell, 2003). One possibility is that RNG impacts upon performance because of this task switching requirement. In Experiment 5 we manipulate instructions in a blocked design in order to evaluate the influence of RNG under belief and logic instruction where there is no requirement to switch between response types. A blocked design also allows an evaluation of performance on the secondary task as a function of the primary task requirements. This is important given that participants will often sacrifice performance on one task in favour of maintaining performance on the other (Gilhooly, Logie, Wetherick & Wynn, 1993; Gilhooly, Logie & Wynn, 2002; Phillips, 1999). One possibility is that the trade-off between the primary and secondary tasks differs as a function of the specific instructional requirements.

3.3 Experiment 5

Task *switching* or *shifting* is an executive function defined by our ability to disengage from an irrelevant task and subsequently actively engage in a relevant task (Miyake et al., 2000). Others argue that shifting should be defined as our ability to perform a new operation in the face of proactive interference having performed a different operation on the same stimuli (Allport & Wylie, 2000). Either way, shifting between mental states is considered an important aspect of executive control required for adequate performance on executive tasks such as the Wisconsin Card Sorting Task (Miyake et al., 2000). Furthermore, shifting has shown to sustain considerable temporal costs (Rogers & Monsell, 1995). Consequently, switching between two instructional sets could be impacting on both accuracy and latency scores. Therefore Experiment 5 aims to control for this possibility, by administering a blocked design. This would eliminate switching demands and also allow us to interpret performance on the secondary task by analysing the randomness data for each condition. The design should provide a clearer indication of the impact RNG has on reasoning performance and allow us to examine the impact of logic and belief judgments on RNG performance.

3.3.1 Method

Participants

Fifty one females and 20 male participants took part, in exchange for course credits.

Design & Materials

A 2 x 2 x 2 x 2 mixed blocked design was used where participants were randomly assigned to belief-first or logic-first blocks within the RNG or Control condition. Each participant received the 128 reasoning problems from Experiment 4, both MP and disjunctive arguments (Complexity) under both belief and logic instructions (Instruction type), which consisted of both conflict and no-conflict problems (Problem type).

Reasoning Task:

A blocked design was used, whereby the 128 problems from Experiment 4, were separated into 4 blocks of 32 arguments. Each block consisted of 16 MP arguments and 16 disjunctives; in two of the blocks participants received belief instructions only and two of the blocks logic instructions only. For a belief block, 16 conclusions required a *believable* response and 16 an *unbelievable* response. For a logic block, 16 conclusions required a *valid* response and 16 an *invalid* response. Each participant was presented with two belief blocks and two logic blocks but they were randomly allocated to belief-first or logic-first for both RNG and Control groups. The belief-first group were presented with block one under belief instruction, block two under logic instruction, block three under belief instruction and block four logic instruction; for the logic-first groups, the reverse applied.

Random Number Generation

The *RGCalc* program was used to analyse the randomness of the numbers generated by each participant (see Experiment 2 & 3, Chapter 2 for more details). Since we were measuring randomness and articulation speed as a dependant measure, we did not exclude participants for being too slow or for poor randomness scores.

Procedure

Before starting the reasoning task, each participant was required to generate a baseline set of random numbers for a total of five minutes. These sets of numbers provided a baseline measure of the randomness indices that could be compared against each experimental block. The control groups were also asked to provide a baseline measure but only as a way of ensuring both conditions (RNG and Control) started the reasoning questions at the same level of cognitive fatigue following the five minute task. The five minute interval was based on previous analysis demonstrating that without a secondary load, participants would take an average of five minutes to complete 32 questions, which was the number of problems allocated to each experimental block.

The instructions were altered to incorporate the baseline measure for each participant. A five minute timer was built into the computer program which would automatically instruct the participants to stop counting when the time was up and proceed to the next part of the experiment.

3.3.2 Results

As in Experiment 4, we eliminated those that scored 50% or below on the conflict items for both the RNG and Control condition. A total of 10 participants were eliminated, six from the RNG group and four from the Control group, which gave a total of N = 30 in the RNG condition and N = 31 in the Control condition.

Randomness Data

We looked at the overall randomness indices across the three` primary task' conditions (belief instruction, logic instruction and baseline). To compare the three independent variables of randomness a MANOVA was carried out to control for type 1 error. Results showed a significant difference in randomness indices based on the primary task; $F(6, 208) = 5.386$, $p < .001$, $\eta_p^2 = .136$. Separate ANOVAs indicated that the primary task had a statistically significant effect on Redundancy scores; $F(2, 105) = 17.316$, $p < .001$, $\eta_p^2 = .248$, but not on Adjacency scores; $F(2, 105) = .458$, $p = .634$, $\eta_p^2 = .009$ or RNG scores; $F(2, 105) = .024$, $p = .976$, $\eta_p^2 < .001$. Post hoc analysis showed that the significant difference for Redundancy scores was between baseline and logic instruction ($p < .001$) and baseline and belief instruction ($p < .001$) but not between belief and logic instruction ($p = .673$). The Adjacency and RNG scores were not significantly different across the three primary tasks. Therefore, the randomness data demonstrated RNG performance was equivalent across the two instructional conditions and that the primary task impacted upon performance on the secondary task.

Accuracy Data

Overall analysis; Block 1 to 4:

A mixed design ANOVA on Arcsine transformed accuracy scores uncovered a main effect of *instruction* (Belief/Logic); $F(1, 59) = 14.359$, $p < .001$, $\eta_p^2 = .196$, confirming that performance was better under logic instruction *(88% vs. 82%).* There was a main effect of *problem* type (Conflict/No-Conflict); $F(1, 59) = 40.138$, $p < .001$, $\eta_p^2 = .405$, showing poorer performance on conflict items *(80% vs. 90%).* There was a main effect of *complexity* (Disjunctives/MP); $F(1, 59) = 22.195$, $p < .001$, $\eta_p^2 = .273$, with MP producing higher accuracy rates *(87% vs. 83%)* and a main effect of *condition*; $F(2, 59) = 8.525$, $p = .005$, $\eta_p^2 = .126$, with RNG reducing accuracy scores *(81% vs. 89%)*: replicating all the main effects of Experiment 4.

Table 3.5

Block 1 – 4: Average Accuracy and Latency scores, across both groups, for Belief and Logic Instructions in each Problem Type, for MP and Disjunctive arguments. Results exclude below chance scores and include correct only latencies. Experiment 5: RNG (N = 30), Control (N = 31).

Variable	Belief Instructions		Belief Instructions		Logic Instructions		Logic Instructions		Over all Means
	MP-c	MP-nc	D-c	D-nc	MP-c	MP-nc	D-c	D-nc	
RNG									
Response Accuracy (%)	74	87	70	84	82	90	77	80	81
Latency Score (ms)	5,460	5,191	6,858	6,743	5,644	6,408	7,672	8,185	6,520
Control									
Response Accuracy (%)	78	94	75	92	91	97	90	95	89
Latency Score (ms)	4,263	3,962	5,841	5,350	4,511	4,263	5,775	5,596	4,945
Mean Accuracy (%) (across each cell)	76	91	73	88	87	94	84	88	
Mean Latency (ms) (across each cell)	4,861	4,577	6,350	6,047	5,078	5,336	6,724	6,891	

There was also an interaction between *instruction* and *problem* type; $F(2, 59) = 9.778$, $p = .003$, $\eta_p^2 = .142$, again demonstrating how the belief/logic conflict had more of an impact on judgments concerning the believability of the conclusion *(74 % B-conflict vs. 90 % B-no-conflict)* rather than its validity *(85% L-conflict vs. 91% L-no-conflict)*.

However, RNG had a larger impact on logic based judgments *(82% RNG vs. 93% control)* compared to belief based judgments *(79% RNG vs. 85% control)*, as indicated by the *instruction* by *condition* interaction; $F(2, 59) = 7.295$, $p = .009$, $\eta_p^2 = .110$. Follow up analysis revealed that the effect was only significant under logic instruction; $F(2, 59) = 15.874$, $p < .001$, $\eta_p^2 = .212$, compared to belief instruction; $F(2, 59) = 1.981$, $p = .165$, $\eta_p^2 = .032$. In order to eliminate practice effects as a possible explanation for the absence of load impact on belief judgments, we carried out separate analysis on both instructions for earlier (1 & 2) and later blocks (3 & 4). Results revealed a marginal effect of RNG on the earlier blocks; $F(2, 59) = 2.943$, $p = .092$, $\eta_p^2 = .048$ and no effect in the later blocks; $F(2, 59) = 2.363$, $p = .130$, $\eta_p^2 = .039$, for belief based judgments, whereas the effect of RNG was present under logic instruction in both the earlier block; $F(2, 59) = 12.120$, $p = .001$, $\eta_p^2 = .170$ and later blocks; $F(2, 59) = 13.722$, $p < .001$, $\eta_p^2 = .189$. This suggests that the absence of an effect of load on belief judgments cannot be explained by belief judgments becoming more automatic over time, thus confirming that RNG has its greatest impact on logic judgments.

Latency Data

Overall analysis; Block 1 to 4:

Analysis carried out on correct only response latencies produced a main effect of *instruction* (Belief/Logic); $F(1, 59) = 5.786$, $p = .019$, $\eta_p^2 = .089$, indicating that belief based judgments took less time to complete than logic based judgments *(5,458ms vs. 6,007ms)*, in contrast to Experiment 4. There was a main effect of *complexity* (Disjunctives/MP); $F(1, 59) = 108.661$, $p < .001$, $\eta_p^2 = .648$, with disjunctives taking longer than MP judgments *(6,502ms vs. 4,962ms)* and a main effect of *condition*; $F(2, 59) = 16.596$, $p < .001$, $\eta_p^2 = .220$, with longer latencies in the RNG group compared to the control group *(6,520ms vs. 4,944ms)*. There was also an interaction between *problem* type and *condition*; $F(1, 59) = 5.170$, $p = .027$, $\eta_p^2 = .081$, which further analysis revealed was driven by the control condition; $F(1, 30) = 5.522$, $p = .026$, $\eta_p^2 = .155$, showing that conflict items took longer to complete *(5,098ms)* compared to no-conflict items *(4,793ms)*. The difference was not significant in the RNG condition; $F(1, 29) = 2.514$ $p = .124$, $\eta_p^2 = .080$ *(6,409ms conflict vs. 6,632ms no-conflict)*. There were no other interactions to report (all *p*'s > .05) (for full data sets see Appendix C).

3.3.3 Discussion

In Experiment 5 a blocked design was adopted in order to minimise the impact of switching demands on task performance. The idea was to evaluate the impact of RNG on the primary task independent of the requirement for participants to switch instructional set as a function of the response options presented. Interestingly, in line with Handley et al. (2011), the impact of conflict was greatest under belief instruction, suggesting that even when participants are aware from trial to trial that they only need to respond based upon the believability of the conclusion; the competing logical structure continues to interfere with their ability to do so. Surprisingly, given the greater difficulty of responding based upon beliefs, the impact of RNG was greatest under logical instruction. This did not arise because participants were differentially allocating resources across the primary and secondary tasks. Although redundancy was greater in the random sequences generated under secondary task conditions, there was no evidence that this differed across instructional conditions. One other important observation in Experiment 5 was that belief judgments were significantly quicker than logic judgments, in contrast to Experiments 1 and 2 (Chapter 2) where trials were interspersed.

Up to this point we have reliably established that people are less accurate under belief instruction, however, in the current experiment belief judgments took less time than in previous studies and in Experiment 4 logic judgments were linked to the AH4 and CRT. The evidence appears to be consistent with the idea that logic judgments are in fact more effortful and more demanding of cognitive capacity. The question is whether there could be an alternative explanation for the lower accuracy scores on belief-conflict problems? One possible explanation for the impact of logic on beliefs concerns the nature of content in the problems presented and participant’s uncertainty about the believability of the conclusions when making belief judgments. Therefore, in order to ensure that uncertainty regarding the belief status of the conclusions is not contributing to the error rates under belief instruction, Experiment 6 will assess the level of consistency in participants evaluations of the believability of the conclusions employed in Experiments 4 and 5.

3.4 Experiment 6

The aim of the current experiment is establish whether there are any discrepancies in the evaluation of conclusions with regards to the believability of its content, for MP and disjunctive arguments.

3.4.1 Method

Participants

Nine females and seven male participants volunteered to take part in this short study, which took approximately five minutes to complete.

3.4.2 Materials and Procedure

Participants were presented with lists of stimuli consisting of information from the 32 MP and 32 Disjunctive arguments used in Experiments 4 & 5. For example; for MP arguments the believability of the conclusion is a function of the believability of the conditional statement. Therefore, in order to check whether participants were consistent in judgments of believability we presented the antecedent of the conditional as a supposition (e.g. suppose a hamster is fed) followed by the consequent as a conclusion (e.g. will the hamster live?) and instructed participants to judge whether the conclusion was believable or unbelievable considering the supposition. From the 32 disjunctive problems participants were presented with a list of concluding statements only (e.g.

does it follow that the sun is green?) and asked whether they thought the information presented in the statement was believable or unbelievable (see Appendix B).

3.4.3 Results

Modus Ponens:

Based on the 16 participants that took part in this experiment, there were a total of 512 judgments for the conditional items presented. Half of the conclusions were used as believable items and half were used as unbelievable items.

32 items x 16 participants = 512 responses

Of the 256 unbelievable conclusions 252 were evaluated as unbelievable and the full 256 believable conclusions were judged to be believable. Overall, this indicates that *99.2%* of judgments were answered in line with the classification of belief status of the conclusion used to classify responses in the experimental work.

Disjunctives:

For the more complex inferences, 249 of the 256 unbelievable conclusions were evaluated as unbelievable and 254 out of 256 were judged to be believable. The results show that *98%* of the judgments were answered according to the responses expected in the experimental trials.

3.4.4 Discussion

Experiment 6 successfully demonstrates a high degree of consistency *(98 – 99%)* in belief judgments and based on these findings it appears unlikely that the results produced so far can be explained by uncertainty about the belief status of the conclusion.

3.5 General Discussion

The principle objective in Experiment 4 was to discount two possible explanations for greater accuracy scores on logic judgments compared to belief judgments. One explanation concerned a possible shortcut strategy with Modus Ponens that would allow participants to generate a logical response by matching the content of the first premise with the conclusion. A second potential explanation for the impact of logical validity on belief judgments is that Modus Ponens inferences are drawn automatically and hence a conclusion is available early that then interferes with belief judgments (Reverberi, Burigo & Cherubini, 2009; Reverberi, Pischedda, Burigo & Cherubini, 2012; Rips, 1988; 1994).

According to Reverberi et al., (2012) disjunctive arguments are not drawn automatically and no shortcut strategy can be used to complete these inferences. Consequently these more complex problems were introduced in Experiment 4 to evaluate these two possible explanations for the findings. Moreover, we were interested in whether RNG would interfere to a greater extent with these more complex logical arguments. The second objective in this chapter was to determine whether switching demands were responsible for the main effect of RNG in Experiment 3 and 4 and if performance on the secondary task would differ under belief and logic instruction.

Experiments 4 and 5 demonstrated, contrary to the DI Dual Process account, that belief based judgments produce more errors than logic based judgments and that conflict between belief and logic instruction results in a greater impact on judging the believability of a conclusion. There was no evidence of a differential impact of logic on belief based judgments between MP and disjunctive arguments. Whilst some research has shown that propositional inferences are made spontaneously even when unnecessary for text comprehension (Lea. 1995), the current findings suggest that the automaticity of the argument does not explain enhanced logic based accuracy in Experiment 3. Experiment 4 used the same method as Experiment 3, and showed that RNG affects both belief and logic judgments, suggesting that both depend upon effortful processing. Whilst RNG did have a greater impact on the more complex disjunctive arguments, the impact was comparable across both instructional sets.

Experiment 5 employed a blocked design to eliminate the possible switching demands on the reasoning task, which could have made judgments more challenging. Switching is thought to involve executive processing, requiring the activation of a new task set, the inhibition of a previous task set, together with the reconfiguration of task set processes (Monsell, 2003). Interestingly, the principle findings regarding accuracy were replicated in the blocked design, but RNG only impacted significantly on logic judgments. The blocked design also enabled us to examine the impact of the primary task on secondary task performance in order to determine whether the effect would differ according to instruction type. The results provided no evidence for this and despite RNG having a greater impact on logic judgments, the generation of random numbers did not differentiate significantly between belief and logic instruction.

Most of the literature suggests that logic based judgments are effortful and belief judgments are less so; however the behavioural data we have presented suggests otherwise. We conjecture in accordance with the PC model (Handley et al., 2011), both belief and logic based judgments are activated early. The underlying structure of the inference is processed rapidly and a logical response is available to the reasoner earlier than a belief based response which requires activation of relevant knowledge and its integration with the conclusion. We would argue that in order to answer according to beliefs, a readily available logical response requires inhibition, which explains why conflict between belief and logic consistently results in a greater impact on judgments of conclusion believability.

How might we reconcile the behavioural data, which suggests belief judgments are most demanding, with the secondary task data, which shows greatest interference with logic judgments? In our view the data were consistent with a Type2/Type2 conflict which arises because of competing responses. The logical response is available early, but nevertheless requires participants to effortfully reason from the underlying structure of the presented argument. This claim is consistent with research that links reasoning from simple propositional arguments with measures of Cognitive Ability (see, for example, Newstead, et al, 2004). In contrast the belief based response is available later and successful responding depends upon the inhibition of an available and competing response. The critical feature of this account is that both types of judgment depend upon Type 2 processing, but rely upon distinct executive resources. A key characteristic of the RNG task is that the successful generation of random numbers depends upon minimising structure in the sequence of responses generated. The task also requires updating of working memory to ensure that generated sequences do not repeat previous ones. It is our conjecture that RNG consequently interferes with the ability to extract the underlying structure of an argument required for processing the validity of an inference but does not impact upon the capacity to inhibit the logical response; hence the larger effect of RNG on logic based judgments in Experiment 5.

In order to explain the shift from the main effect of RNG (Experiments 3 and 4) to its increased impact on logic judgments, we have assumed that the switching between instructional sets in a mixed design (disengaging from one irrelevant task and effectively engaging in a relevant one) places additional demands on executive resources. These switching demands combined with RNG increase the executive demands of the task consequently impacting upon both types of judgment. More specifically, switching between belief and logic requires the ability to inhibit one instructional set for another; therefore the capacity to supress a fast completing logic based response is affected from the offset, even if RNG does not impact on the capability of inhibiting a logical output. In a blocked design the inhibition required to switch instructional sets is eliminated, therefore belief based judgments are less affected. However, logical outputs still require inhibition, hence the belief/logic conflict. Random number generation nevertheless is still demanding of executive resources thus affecting logic based inferences and offering an explanation for the overall effect of load in a mixed design.

As argued above, this account implicates effortful processing in both judgments types and a distinction in essential executive resources. This proposal receives some support from the results of the AH4, and CRT. In Experiment 4 participants with high AH4 scores show higher accuracy on logical judgments than low AH4 scorers whilst the difference in accuracy on belief judgments is smaller. Although the effects are small in size, a similar pattern is observed with the CRT data suggesting that higher cognitive ability linked to the AH4 and cognitive style as measured by the CRT are associated with greater sensitivity to logical structure. Traditionally, the CRT is a measure of cognitive style designed to assess the inclination to override prepotent, incorrect responses by employing further reflection. We propose that the CRT also relates to a person's ability to extract the underlying structure of the problem in order to arrive at the correct conclusion; hence its association with validity judgments. Interestingly, the data do not show that the CRT is linked to less belief-bias under logical instructions.

Perhaps our findings suggest a Type 2/Type 2 conflict but that belief and logic depend on different executive resources. We attempted to address this notion by examining the randomness indices produced under each instructional set. Different indices generated by the *RGCalc* program are supposed to measure the distinct executive functions involved in RNG (Miyake et al., 2000), however, there was no differential allocation of resources to the secondary task as a function of instruction. Perhaps randomness indices are not a robust enough measure of executive functions; therefore the next chapter will focus on evaluating the separate executive functions of inhibition and working memory in order to determine whether belief and logic judgments pull on distinct resources as explanation for the differential effect of the secondary task displayed in this chapter.

Chapter 4: Executive Resources and Reasoning

4.1 Introduction

In Chapter 3 we established that secondary load impacted on logic judgments more so than belief judgments but the impact of conflict was greatest under belief instruction, both when problems were interspersed (in line with previous findings) and when presented in a blocked design. How then do we integrate these two opposing results? One possible interpretation for these findings is that each type of judgment depends on distinct executive resources. The experimental data presented in the previous chapter supports the idea that logical responses to MP and disjunctives arguments are available early and require *inhibiting* in order to generate belief based responses. However, the greater impact of the load on logic judgments also suggests some effortful Type 2 processing required for judgments based on the validity of an argument. We conjecture that RNG impacts on logical judgments by interfering with the ability to reason from the structure of the argument but does not affect the ability to inhibit a Type 1 logical output.

Miyake et al. (2000) specified both the role of `inhibition' and ` updating' in RNG, however some researchers suggests that generating random numbers predominantly involves supressing stereotyped responses (Baddeley, 1998) whilst others posit that the main component is keeping track of recent responses and comparing them to a representation of what is considered random (Jahanshahi et al., 1998). Earlier we stated that RNG does not seem to impact on the ability to *inhibit* a logical response but conjecture that it affects the ability to monitor and update working memory representations as an explanation of the effect of load on responding to the logical structure of an argument. Consequently, we still assume that judgments based on the believability of an argument pull on the executive function of inhibition, whereas logical inferences demand WM for the representation and integration of premise information. Therefore, the principle aim of the current chapter is to examine the notion that distinct executive resources are required for belief and logic judgments, by investigating the role of WM with a memory *updating* *n*-back task and the role of *inhibition* with an original take on the Stroop task.

We predict that responses based on the logical structure of the argument will be most affected by the *n*-back task, and the Stroop task will have its greatest impact on belief based judgments.

4.2 Experiment 7

The objective of the current experiment is to examine the function of WM in relation to belief and logic based judgments, with the intention of explaining the conflict that arises because of competing responses potentially engaging different executive processes.

In Experiment 7 we will investigate the role of memory updating in belief and logic based judgments. Memory updating is the executive function that monitors incoming information for its magnitude and appropriateness and updates the information in WM by exchanging old for new (Morris & Jones, 1990). The present experiment will add an additional instructional condition to the reasoning task that involves the identification of features presented in a previous trial (see Design and Materials section for more details). This *n*-back design means participants will be unable to predict whether they will be required to judge the validity or believability of a conclusion or recall the features in the preceding trial; thus increasing the demands of WM.

The N-Back Task

N-back is based on a task created by Wayne Kirchner in 1958 to determine age differences in the short term retention for changing information (Kirchner, 1958). The *n*-back is a performance task requiring continuous maintaining, updating, and releasing of arbitrary bindings between items and temporal order positions (Friedman et al., 2006). It is generally used to measure a part of WM and has been proposed as a technique for increasing fluid intelligence (Gf). It has also been a useful tool in the research on cognitive aging (Kirchner, 1958; Schmiedek, Li & Lindenberger, 2009), schizophrenia (Glahn et al., 2005) and has been used in cognitive neurosciences to look for commonalities and differences in brain activation produced by manipulations of the process and content of WM (Owen, McMillan, Laird & Bullmore, 2005).

The task involves showing participants a sequence of letters or numbers and asking them to evaluate whether the current stimulus matches one presented earlier in the sequence or *n*- back, where *n* is commonly equal to one, two or three. For example, in a three back task, participants would have to say for each stimulus whether it matched the one presented three steps back. The task requires keeping the relevant information in WM and updating the contents as each new piece of information is provided, whilst n+1 back is discarded.

This experiment will use an original variant of the *n*-back task (see Materials) to explore the role of WM in the logic and belief judgment task. Additionally, we will employ a range of individual difference measures as a supplementary assessment of executive functioning, with the use of an Operation Span Task (adapted from Turner & Engle, 1989), a short Stroop Task (Stroop, 1935) and the Cognitive Reflection Task (Frederick, 2005) in order to reveal any relationships between these measures and task performance.

4.2.1 Method

Participants

A total of 81 participants took part in exchange for two course credits. Seventy six females and five male participants were randomly assigned to either the Experimental condition (Memory Updating) or Control condition (No Updating).

Design

This experiment integrated the *n*-back task into the reasoning task, creating a 3 x 2 x 2 x 2 mixed design. Participants were required to make judgments on the validity or believability of a conclusion or on the font style of the last word in the previous trial, creating three *instructional* sets and increasing the demands of WM in the experimental condition. Participants were presented with both conflict and no-conflict problems (*problem type*) in the form of MP and disjunctive arguments (*complexity*) in either the Experimental or Control group (*condition*) as the between subject factor.

Jonides and Smith, (1997) argued that response inhibition can also play a role in the *n*-back task, for example, in a three back task, there may be occasions when two back or one back matches the current stimulus, which would require the inhibition of the natural tendency to respond based on matching rather than updating. Therefore, in order to minimise the role of inhibition in the task, we ran a 1-back task.

Material

Reasoning Task:

A novel set of 192 conditional arguments were created for Experiment 7. Each participant received 64 belief trials, 64 logic trials and 64 *n*-back trials. The trials in each condition consisted of 32 MP and 32 Disjunctives arguments. For the belief and logic trials, 16 of each set were conflict problems, with valid/unbelievable or invalid/believable conclusions and the remaining 16 were no-conflict problems with conclusions that were valid/believable and invalid/unbelievable (see Appendix B).

Experimental Manipulation - N-back Task:

The 64 *n*-back trials were administered to each participant in both the Experimental and Control condition. In both conditions the font style of the last word in the concluding sentence was altered and participants were required to indicate whether the font of the last word in the previous trial matched the font of the current trial by responding `same' or `different' (see Table 4.1). The following five font styles were used in size 18 font: Lucida Handwriting, Bradley Hand ITC, ALGERIAN, Ravie and Curlz MT.

Table 4.1
Examples of the integrated reasoning/n-back task and the format in which the trials were presented for both the Experimental and Control conditions.

Previous Trial		**Current Trial**	
Either Flamingos are pink or they are purple		If the bird is a Dove then it is orange	
Suppose flamingos are pink		Suppose the bird is a Dove	
Does it follow that flamingos are purple?		Does it follow that the bird is WHITE?	
s) believable	k) **unbelievable**	s) same	k) **different**

The *n*-back trials were incorporated into the reasoning task, adding a third instructional set to the design. In the Experimental condition we increased the demands on WM by having an equal number of `same' and `different' responses which required participants to update memory from trial to trial. A total of 32 *n*-back trials were *different* to those `1-back' and 32 were the *same* and these were evenly distributed across MP and disjunctive arguments. Both conflict and no-conflict problems were also assigned equally often to both response options and conflict trials were either unbelievable or invalid whilst no-conflict trials were believable and valid.

In the Control condition the 64 *n*-back trials were also assigned equally often to the conflict and no-conflict problems but there was no requirement to update working memory because the font style always matched the font in the previous trial. In other words, the trials only ever required a `same' response in the control condition.

The *n*-back trials were interspersed with the reasoning questions but had to be specifically placed in positions were the previous trial was either the `same' or `different'. For this reason, we created 32 distinct pseudo-random orders in which to present the 192 trials.

Measures of Individual Differences:

Short Stroop Task

The Short Stroop task was taken from a paper by Raz, Shapiro, Fan & Posner (2002) and adapted for computer administration. The stimuli comprised of a word written in one of four ink colours (yellow, blue, red and green) appearing at the centre of the computer screen where a black fixation cross was previously located. Two classes of words were used: colour words (Yellow, Blue, Red and Green) and neutral words (Flower, Ship, Lot and Knife). The neutral words match the coloured words on frequency and length and all characters were displayed on a white background.

The task consisted of six practice trials with feedback to begin, followed by 36 experimental trials. These consist of 12 neutral trials, where the neutral word was in any of the four colours, 12 congruent trials where the colour word matched the colour and 12 incongruent trials where the colour word was in any of the three colours other than the one to which it referred (e.g., the colour word BLUE in green ink).

Participants were instructed to indicate the ink colour of the word by pressing one of the following four keys; V, B, N, and M, which correspond to the colours red, blue, green, and yellow. Two fingers of each hand were used to press these response keys (e.g., left middle finger for V and right index finger for N). Participants were instructed to focus their eyes upon a fixation cross at the centre of the screen until a stimulus appeared replacing the crosshair. The stimulus remained on the screen until participants responded but speed and accuracy were emphasized equally (see Appendix B for materials).

The Stroop Interference Effect (SIE) was calculated by taking the average reaction time (RT) scores for neutral trials from the average RT scores of incongruent trials (I-N) as a measure of inhibition.

Operation Span Task

The Operation Span task (adapted from Turner & Engle, 1989) presented participants with a set of *equation–word* pairs on a computer screen. Each trial consisted of a simple arithmetic operation followed by a single word (e.g., ‘‘bear’’) they had to recall at the end of the set.

For example:

(3 x 4) + 11 = 22

Bear

Participants had to determine whether the sum of the equation given was correct or incorrect by clicking on `Yes' for correct or `No' for incorrect. They then had to memorise the word presented after the equation.

Each equation remained onscreen until a verification response was given and the next trial was presented or for a maximum of 8 seconds. Once the equation disappeared, the word was presented for 750ms before the next equation was displayed. At the end of the set of trials, they had to recall all the words by typing them into the blank boxes displayed on the screen, in the same order in which they were presented.

The task commenced with three practice trials followed by three target trials at each set size, which ranged from two to five. In other words, there were three trials at set size two (two equations per trial) and at the end of each of the two equation trials, participants had two words to recall. In set size three there were three trials with three equations in each trial and at the end of each trial there were three words to recall and so on to the maximum of five, which gave a total of 42 words to be recalled.

Participants were told they were measured on speed as well as accuracy ensuring they solved the equations rapidly and helped avert possible rehearsal strategies.

Global Span (GS) scores were calculated by totalling the number of correct word trials in each set. For example, a two word set had three trials of two words to remember in each trial, giving a top score of six points. A three word set had three trials of three words to remember, with a top score of nine points to reach. In order to obtain full points for a trial, every word had to be recalled correctly in the right order, or a score of zero was given for that trial. The GS score was a running total of points from each set which was then used in a correlational analysis with reasoning performance (See Appendix B for materials).

Cognitive Reflection Task

The CRT consisted of the three questions presented in Experiment 4, Chapter 3, which were administered to each participant, post reasoning trials. Results for the CRT were calculated by separating participants into high and low groups. The high group comprised of those that scored one or more out of three whilst the low group failed to give any correct answers.

Procedure

Participants were tested in maximum groups of four in the same partitioned booths described in Chapter 2 and were randomly assigned to the experimental or control group. In both conditions, each participant was presented with 128 reasoning questions (64MP and 64 Disjunctives) plus the 64 *n*-back trials with an optional respite at the half way point. The experiment was presented on a computer screen, through an E-prime program. The instructions for the reasoning task were the same as those presented in Experiment 4 (Chapter 3), with additional instruction for the *n*-back trials. Twelve practice trials with feedback were administered to each participant prior to the main part of the experiment. Instructions for the *n*-back trials highlighted the importance of keeping track of the font in the previous trial since they would remain naïve to response requirements (i.e. whether they should answer according to logic, belief or font style) until the response options were presented. The same *n*-back instructions were given in the control condition, even though the *n*-back trials always matched the font style in the previous trial.

On completion of the 192 trials, a short Stroop Task and Operation Span Task were administered via computer, followed by the Cognitive Reflection Task which was completed on paper.

4.2.2 Results

As in previous studies, we eliminated participants who scored 50% or below on the conflict items for both the Experimental and Control conditions. A total of six participants were eliminated, three from the Experimental condition (N = 37) and three from the Control condition (N = 38). Analyses were carried out on accuracy measures that were Arcsine transformed and on correct responses for the latencies. Outliers were removed from the latency data by two standard deviations from the mean of each cell and any missing data was replaced with the overall cell mean. In all of the experiments reported in this chapter, missing data accounted for no more than 5.5% of the overall data. The tables throughout this chapter present response latencies and percentage accuracy scores, prior to transformation.

The first set of analyses presented here is for the *n*-back instruction alone, followed by a second set of analyses evaluating the impact of belief and logic instruction on reasoning performance. The reason for examining the data in this way was to help us evaluate the effectiveness of the experimental manipulation. In terms of *n*-back performance we expected that accuracy would be influenced by the complexity of the argument and the presence of conflict. With regards to belief and logic instruction we expected the additional working memory demands of the *n*-back task to impact upon accuracy on the reasoning task. Specifically we predicted that logic judgments would be affected by the requirement to remember characteristics of the previous trial.

N-Back Instruction

A 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(Experimental/Control Condition) mixed design ANOVA was carried out on Arcsine transformed accuracy data under *n*-back instruction. The *problem* type and *complexity* variables presented here, reflect the type of problem that accompanies the *n*-back trials. The results revealed no influence of *problem* type on *n*-back accuracy (Conflict/No-Conflict); $F(1, 73) = 2.243$, $p = .139$, $\eta_p^2 = .030$, but a significant main effect of *complexity* (Disjunctives/MP); $F(1, 73) = 4.094$, $p = .047$, $\eta_p^2 = .053$, where *n*-back trials accompanying disjunctives resulted in lower accuracy scores than those accompanying MP *(78% vs. 83%)*. This suggested that the more complex the problem, the more challenging it was for participants to recall the font style of the previous trial. Surprisingly, there was no main effect of *condition;* $F(2, 73) = 1.908$, $p = .171$, $\eta_p^2 = .025$, showing no significant difference in performance between the control and experimental condition *(78% vs. 82%)*. In fact the means suggest performance was poorer in the control condition, further supported by the interaction between *complexity* and *condition*, $F(1, 73) = 4.105$, $p = .046$, $\eta_p^2 = .053$, demonstrating that *n*-back accuracy on MP judgments was lower than disjunctives judgments in the control condition than experimental condition *(79% MP-Control vs. 86% MP-experimental)*, whereas there was no effect of condition on disjunctive judgments *(78% Disj-Control vs. 79% Disj-experimental)* under *n*-back instruction. Table 4.2 indicates that performance was generally poorer across most cells for the control condition, apart from conflict-disjunctive items which underpins the 3 way interaction between *complexity*, *problem* type and *condition;* $F(1, 73) = 5.257$, $p = .025$, $\eta_p^2 = .067$. As Table 4.2 indicates, this interaction appears to reflect a difference in *n*-back performance between MP and disjunctive-conflict problems in the experimental condition but not in the control condition. However, follow up analysis did not produce any significant interactions between *problem* and *complexity* for either the experimental condition; $F(1, 36) = 2.434$, $p = .127$, $\eta_p^2 = .063$ or the control condition; $F(1, 37) = 3.326$, $p = .076$, $\eta_p^2 = .082$. All remaining interactions were not significant (all $p > .05$).

Table 4.2

Average Accuracy and Latency scores on Conflict and No-conflict Problems for MP and Disjunctives items under n-back instruction, across both Conditions. Results exclude below chance scores and include correct only latencies. Experiment 7: Experimental Condition (N = 37), Control condition (N=38).

Variable	Conflict		No-conflict		Overall means
	MP	Disjunctives	MP	Disjunctives	
Experimental					
Response Accuracy (%)	85	74	86	83	82
Latency (ms)	3,179	3,056	3,319	3,523	3,269
Control					
Response Accuracy (%)	78	78	79	77	78
Latency (ms)	3,083	3,275	3,175	3,302	3,209
Mean Accuracy (%) (across each cell)	82	76	83	80	
Mean Latency (ms) (across each cell)	3,131	3,166	3,247	3,413	

The mixed design ANOVA on response latencies for *n*-back instruction only produced a main effect of *problem* type; $F(1, 73) = 4.746$, $p = .033$, $\eta_p^2 = .061$, with *n*-back judgments associated with no-conflict problems taking longer than conflict problems (*3,337ms vs. 3,148ms*) but again there was no main effect of *condition*; $F(2, 73) = .065$, $p = .800$, $\eta_p^2 = .001$, showing comparable latencies for the experimental and control condition (*3,269ms vs. 3,209ms*). All remaining interactions were not significant (all $p > .05$).

Belief and Logic Instruction

The second set of analyses on the accuracy data was a 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(Experimental/Control Condition) mixed design

ANOVA on both belief and logic instruction. Results produced a main effect of *instruction* (Belief/Logic); $F(1, 73) = 35{,}460$, $p < .001$, $\boldsymbol{\eta_p^2} = .327$, with better performance on logic compared to belief judgments *(94% vs. 89%)* and a main effect of *problem* type (Conflict/No-Conflict); $F(1, 73) = 33.004$, $p < .001$, $\boldsymbol{\eta_p^2} = .311$, with poorer performance on conflict items compared to no-conflict items *(88% vs. 95%).* There was a main effect of *complexity* (Disjunctives/MP); $F(1, 73) = 37.279$, $p < .001$, $\boldsymbol{\eta_p^2} = .338$, where disjunctives produced lower accuracy scores than MP *(89% vs. 94%)* and no main effect of *condition*; $F(2, 73) = .683$, $p = .411$, $\boldsymbol{\eta_p^2} = .009$, showing no significant difference in accuracy scores across conditions *(91% experimental vs. 92% control).*

There was a significant interaction between *instruction* and *problem* type; $F(1, 73) = 10.899$, $p = .001$, $\boldsymbol{\eta_p^2} = .130$, consistent with previous findings where conflict between belief and logic had a greater effect on judgments of the believability of a conclusion (*83% B-conflict vs. 94% B-no-conflict)* than judgments of the conclusion's validity *(95% L-conflict vs. 93% L-no-conflict)* (see Table 4.3).

Table 4.3

Average Accuracy and Latency scores for Belief and Logic Instruction on Conflict and No-conflict Problems, for MP and Disjunctive arguments. Results exclude below chance scores and include correct only latencies. Experimental Condition (N = 37), Control condition (N=38)

Variable	Belief Conflict		Belief No-conflict		Logic Conflict		Logic No-conflict		Overall Means
	MP	Disj	MP	Disj	MP	Disj	MP	Disj	
Experimental									
Response Accuracy (%)	87	86	96	89	94	88	98	88	91
Latency (ms)	6,838	6,646	6,178	6,457	7,189	6,950	6,747	6,808	6,727
Control									
Response Accuracy (%)	80	80	98	92	98	92	99	93	92
Latency (ms)	7,151	6,944	6,462	6,599	6,734	7,458	6,669	6,940	6,870
Mean Accuracy (%) (across each cell)	84	83	97	91	96	90	99	91	
Mean Latency (ms) (across each cell)	6,995	6,795	6,320	6,528	6,962	7,204	6,708	6,874	

There was a significant interaction between *problem* type and *complexity*; *F*(1, 73) = 14.367, *p* < .001, η_p^2 = .164, suggesting that conflict had less of an impact on disjunctive judgments *(87% Conflict vs. 91% No-conflict)* than MP judgments *(90% Conflict vs. 98% No-conflict)*. There was an interaction between *instruction* and *complexity*; *F*(1, 73) = 4.586, *p* = .036, η_p^2 = .059, with a bigger difference in performance between MP and disjunctive judgments under logic instruction *(98% L-MP vs. 91% L-Disjunctives)* than under belief instruction *(91% B-MP vs. 87% B-Disjunctives)*. Finally there was a significant interaction between *instruction* and *condition; F*(1, 73) = 4.073, *p* = .047, η_p^2 = .053, showing a bigger effect of condition on logic judgments (*92% experimental vs. 96% control)* than on belief judgments *(90% experimental vs. 88% control)* suggesting that the *n*-back task had a greater impact on logic, although it is worth noting that this effect was small.

The mixed design ANOVA on latencies produced a main effect of *instruction*; $F(1, 73) = 5.628$, $p < .020$, $\eta_p^2 = .072$, which showed logic based judgments taking longer than belief judgments (*6,937ms vs. 6,660ms*). This could suggest that the presence of the *n*-back task slows down logic judgments. There was a main effect of *problem* type*;* $F(1, 73) = 14.403$, $p < .001$, $\eta_p^2 = .165$, with conflict items taking longer than no-conflict items (*6,989ms vs. 6,608ms*) but there was no main effect of *condition*; $F(2, 73) = .153$, $p = .697$, $\eta_p^2 = .002$, showing no significant difference in latencies between the conditions. All remaining interactions were not significant (all $p > .10$) (see Appendix C for full data sets).

The next set of analyses reports the main effects and interactions between each experimental variable and the CRT measure. Participants were assigned to high and low CRT groups and a 2 (Belief/Logic) x 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(High and Low CRT) mixed design ANOVA was carried out on all accuracy scores collapsing across both conditions. The results revealed a main effect of CRT; $F(2, 73) = 7.462$, $p = .008$, $\eta_p^2 = .093$, where high CRT scorers produced better accuracy scores *(95%)* than low CRT scorers *(89%).* There were also significant 3 way interactions between *instruction*, *problem* and *CRT*; $F(1, 73) = 4.683$, $p = .034$, $\eta_p^2 = .060$, and between *problem* type, *complexity* and *CRT*; $F(1, 73) = 4.151$, $p = .045$, $\eta_p^2 = .054$. Separate analysis on the High (N = 26) and Low (N = 49) groups revealed that the interaction between *instruction* and *problem* type was significant in the low CRT group; $F(1, 48) = 12.431$, $p = .001$, $\eta_p^2 = .206$, but not significant in the high group; $F(1, 25) = .110$, $p = .743$, $\eta_p^2 = .004$. This shows that conflict has a greater impact on belief judgments *(79% conflict vs. 93% no-conflict)* than logic judgments *(92% conflict vs. 93% no-conflict)* for those in the low CRT group, compared to belief judgments *(92% conflict vs. 96% no-conflict)* and logic judgments *(96% conflict vs. 98% no-conflict)* in the high CRT group. Similarly, the interaction between *problem* type and *complexity* was carried by the low CRT group alone; $F(1, 48) = 16.376$, $p < .001$, $\eta_p^2 = .254$, with conflict impacting on MP judgments *(87% conflict vs. 97% no-conflict)* more so than disjunctives *(84% conflict vs. 89% no-conflict)*, whereas the difference between conflict and no-conflict problems was comparable for MP*(95% conflict vs. 99% no-conflict)* and disjunctives *(92% conflict vs. 95% no-conflict)* in the high CRT group; $F(1, 25) = .467$, $p = .500$, $\eta_p^2 = .018$. The latency data produced no main effect or interactions with CRT (all p > .05).

Table 4.4

Average Accuracy scores for Belief and Logic Instructions across Conflict and No-conflict problems for MP and Disjunctive arguments. Results exclude below chance accuracy scores across High and low CRT. High CRT (N = 26), Low CRT (N=49).

Variable	Belief Conflict		Belief No-Conflict		Logic Conflict		Logic No-Conflict		Overall Means
	MP	Disj	MP	Disj	MP	Disj	MP	Disj	
High CRT Average Response (%)	93	90	98	94	97	94	99	96	95
Low CRT Average Response (%)	78	79	96	89	96	88	98	88	89
Mean Accuracy (%) (across each cell)	86	85	97	92	97	91	99	92	

The Global Span (GS) scores produced by each participant were correlated with the reasoning accuracy data across each cell (See Table 4.5). The results produced a significant positive correlation between GS scores and accuracy on belief items; $r(73) = .330$, $p = .004$, specifically belief-conflict items; $r(73) = .315$, $p = .006$, suggesting that higher GS scores result in better performance on belief-conflict problems, which does not extend to logic-conflict problems; $r(73) = .167$, $p = .152$.

The Stroop Interference Effect (SIE) was correlated with reasoning accuracy scores from each cell and results revealed a significant negative correlation with belief trials; $r(73) = -.310$, $p = .007$, specifically belief-conflict items; $r(73) = -.295$, $p = .010$, which suggests that those who experience more Stroop interference perform worse on belief conflict judgments; again these findings do not extend to logic judgments; $r(73) = -.028$, $p = .814$.

Table 4.5
Correlations for Global Span scores and Stoop Interference Effect, measured against each Instructional set and for both Problem Types. Results exclude below chance accuracy scores.

GS = Global Span; SIE = Stroop Interference Effect

	Belief No-Conflict	Belief Conflict	Overall Belief	Logic No-Conflict	Logic Conflict	Overall Logic
GS	.204	.315**	.330**	.171	.167	.177
SIE	-.196	-.295*	-.310**	.043	-.090	-.028

Key: ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

4.2.3 Discussion

The principle aim of Experiment 7 was to investigate whether increasing WM demands of the task by incorporating an additional requirement to remember and update characteristics associated with the previous trial would have a greater impact on logical judgments. The overall main effects from this experiment produced the same outcomes as Experiments 4 & 5, with lower accuracy scores for belief judgment, lower scores when the problems were more complex and when there was conflict between belief and logic. Additionally, the results confirmed that belief/logic conflict had a greater impact on the believability of the conclusion.

In support of our initial prediction, the experimental condition had more of an effect on logic based judgments as indicated by the interaction between *instruction* and *condition* (see analysis on belief and logic instruction). Furthermore, logic judgments took longer to complete than those based on beliefs. Evaluation of the results produced under *n*-back instruction revealed that the complexity of the argument impacted on participants' ability to recall the characteristics of the previous trial, also suggesting that the *n*-back task interfered with processing the structure of the argument. Unexpectedly, however, there was no main effect of condition for accuracy or latency scores under *n*-back instruction. This indicates that the experimental manipulation did not create greater demands on participants than the control condition. In fact the means show (although not statistically significant) that across the majority of cells, participants produced lower accuracy scores in the control condition under *n*-back instruction. The only cells that produced significantly lower accuracy scores in the experimental condition were conflict-disjunctive items, however the fact that there was a main effect of complexity suggests that the structure of the argument was still having an effect on what should have been a simple control task. This indicates that participants found most of the *n*-back trials in the control condition as equally challenging as those in the experimental condition.

There is one possible explanation for comparable accuracy and latency scores across conditions, under *n*-back instruction. Perhaps the control task created an extra memory load as a result of the continuous uncertainty about whether the current trial would ever mismatch the previous trial. For example, in the experimental condition, the correct answer for the *n*-back trials could be either the `same' or `different', whereas in the control condition, the correct response was always the `same'. Unless participants promptly realised this pattern (and some did) they would attempt to keep the font style updated in their WM from trial to trial `just in case' they differed; thereby creating a similar demand on WM resources to the experimental condition.

Analyses on the CRT scores produced some contrary findings to those presented in Experiment 4 which showed that CRT was related to logic based reasoning more so than belief based reasoning. In the present experiment the results revealed that conflict had its greatest impact on belief judgments for the low CRT scorers. One interpretation for this is that those with a lower ability to inhibit an automatic (logical) response will find it harder to resolve belief/logic conflict in favour of a belief based output. However, since we only have two (opposing) data sets to base our interpretation on, we will reserve an explanation of the CRT findings for the general discussion.

The operation span task has been widely used as measure of WM capacity and is acknowledged in requiring the operations of the central executive (Engle, Tuholski, Laughlin & Conway, 1999). In the current experiment we found that GS scores correlated with belief conflict trials suggesting that beliefs draw more heavily on WM resources. Furthermore, those better equipped at inhibiting a conflicting incorrect response, as measured by Stroop Interference, performed better on belief-conflict judgments. Neither of the measures correlated with performance on logic trials, which could be explained by the fact that logical performance was close to ceiling. However, logic judgments did take longer to complete, indicating that the presence of *n*-back trials slowed down logical reasoning without having an effect on accuracy. Again we will discuss these findings in more detail in the general discussion.

The present study appears to confirm our original hypothesis that the *n*-back task has its greatest effect on logic based judgments; however the concern is that the overall accuracy scores between the experimental and control condition did not differ significantly. Consequently, it is difficult to conclude that the experimental manipulation had more of an impact on logic judgments when in effect the control task was functioning in a similar fashion to the experimental task and even proved more challenging for some participants. In other words the interaction between *instruction* and *condition* could instead be interpreted as the control task having more of an impact on belief based judgments *(88% belief vs. 96% logic)* compared to the experimental condition *(90% belief vs. 92% logic)*. Nevertheless, we reserve any further conjecture about the current findings until we can ensure that the control task is not inadvertently exerting similar (or more) demands on WM than the experimental condition. Therefore, in order to minimise the demands of the control task, Experiment 8 uses a matching task for the control condition.

4.3 Experiment 8

In the current experiment we will use the same methodology for the Experimental condition as Experiment 7, integrating the *n*-back task into the reasoning task and asking participants to make judgments based on belief, logic or the font of the last word in a preceding trial. For the Control condition participants will be required to *match* the font style of the last word in the major premise with the font of the last word in the concluding sentence of the same trial; thereby eliminating any need to store and update information based upon the characteristics of the previous trial. This time we expect a significant difference between the conditions under *n*-back instruction which will allow us to more clearly evaluate the impact of a memory load on the judgment task relative to an undemanding control condition.

4.3.1 Method

Participants

A total of 79 participants took part in the current study, in exchange for course credits or £8 payment. Fifty nine females and 20 male participants were randomly assigned to either the Experimental (Memory Updating) or Control condition (Matching).

Design & Material

Experimental Condition:

The design for the Experimental condition was the same as Experiment 7 with the *n*-back trials integrated into the reasoning task, creating a 3 x 2 x 2 x 2 mixed design. The same 64 Modus Ponens (MP), 64 Disjunctives and 64 *n*-back trials were used with the three types of *instruction* asking participants to make judgments based on conclusion believability, conclusion validity or the font style of the last word in the previous trial. Again participants were presented with both conflict and no-conflict problems in the form of MP and disjunctive arguments for the reasoning task in both the Experimental and Control conditions (refer to Reasoning Task in Experiment 7).

Control Condition:

In the Control condition the 64 *n*-back trials were converted into 64 matching trials where participants were required to *match* the font style of the last word in premise one with the font style of the last word in the conclusion (see Table 4.5). Similar to the *n*-back trials 32 of the trials matched in font style and required a `same' response and the 32 mismatched, requiring a `different' response. An equal number of MP and Disjunctive arguments were distributed to matching and mismatching trials of which conflict and no-conflict problems were also evenly distributed to both response options.

Participants were presented with the same 128 reasoning trials as the Experimental condition, interspersed with 64 matching trials. As in Experiment 7, the full 192 trials were again presented in one of the 16 unique sequential orders created, to guarantee pseudo-randomisation of the stimuli.

Table 4.6
Examples of the Matching Trials presented in the Control Condition.

Matching Trial		**Mismatching Trial**	
Either the sun is yellow or it is blue		If the bird is a dove then it is orange	
Suppose the sun is blue		Suppose the bird is a dove	
Does it follow that the sun is yellow?		Does it follow that the bird is WHITE?	
s) **same**	k) different	s) same	k) **different**

Measures of Individual Differences:

The **Short Stroop Task** and **Operation Span Task** were completed after the reasoning task, followed by the **CRT (**refer to Experiments 7 for details).

Procedure

The procedure was equivalent to Experiment 7 for both the Experimental and Control conditions, with participants being tested in maximum groups of 4, randomly allocated to the control or experimental group. In both conditions participants were presented with the 128 reasoning questions and 64 *n*-back trials as the experimental manipulation or 64 matching trials as the control task. Each group was presented with 12 practice trials before the main experiment. The instructions for the *n*-back task emphasised the importance of remembering the characteristics of the previous trial, whilst the control task restricted the matching of font style to the current trial, thereby eliminating any unnecessary WM demands that may have been created by the control condition in the previous experiment.

After the reasoning task, participants completed the short Stroop Task and Operation Span Task on the computer, followed by the written Cognitive Reflection Task (see Experiment 7 for more details).

4.3.2 Results

A total of six participants were eliminated, two from the Experimental condition (N = 37) and four from the Control condition (N = 36), for performing below chance on conflict items. Analogous to Experiment 7, the first set of analyses presented is for the *n*-back and matching instruction, in order to examine the demands of the experimental manipulation compared to the control condition. This is followed by separate analyses on belief and logic instruction.

N-Back vs. Matching Instruction

A 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(Experimental/Control Condition) mixed design ANOVA was carried out on Arcsine transformed accuracy data under *n*-back instruction. The *problem* type and *complexity* variables presented, reflect the type of problem that accompanies both *n*-back and matching trials. The results uncovered a marginal main effect of *problem* type (Conflict/No-Conflict); $F(1, 71) = 3.827$, $p = .054$, $\eta_p^2 = .051$, showing that the *n*-back and matching trials accompanying conflict items produced marginally lower scores than no-conflict items *(83% vs. 86%).* There was also a significant main effect of *complexity* (Disjunctives/MP); $F(1, 71) = 12.712$, $p = .001$, $\eta_p^2 = .152$, where *n*-back and matching trials accompanying disjunctives produced lower accuracy scores than those accompanying MP *(82% vs. 87%).* This suggested that participants found it more challenging to recall or match font style when the problem was more complex. Finally, there was a main effect of *condition; F*(2, 71) = 17.132, *p* < .001, η_p^2 = .194, which confirmed that performance was lower in the experimental condition than the control condition *(80% vs. 89%)*. There were no significant interactions to report (all *p* > .05).

Table 4.7
Average Accuracy and Latency scores on Conflict and No-conflict Problems for MP and Disjunctives items under n-back Instruction (experimental) and Matching Instruction (control). Results exclude below chance scores and include correct only latencies. Experiment 8: Experimental Condition (N = 37), Control condition (N=36).

Variable	Conflict		No-conflict		Overall means
	MP	Disjunctives	MP	Disjunctives	
Experimental					
Response Accuracy (%)	81	76	84	81	81
Latency (ms)	3,265	3,411	3,007	3,324	3,252
Control					
Response Accuracy (%)	92	85	93	88	90
Latency (ms)	3,973	4,504	3,960	4,721	4,290
Mean Accuracy (%) (across each cell)	87	81	89	85	
Mean Latency (ms) (across each cell)	3,619	3,956	3,484	4,023	

Response latencies for *n*-back instruction produced no main effect of *problem* type*; F*(1, 71) = .144, *p* = .705, η_p^2 = .002. There was a main effect of *complexity* type*; F*(1, 71) = 22.106, *p* < .001, η_p^2 = .237, showing that *n*-back and matching trials accompanying disjunctive judgments took longer to complete than MP *(3,990ms-Disjunctives vs. 3,551ms-MP)*. There was also a main effect of *condition*; *F*(2, 71) = 13.691, *p* < .001, η_p^2 = .162, showing that performance in the control condition was significantly slower than in the experimental condition *(4,290ms vs. 3,252ms)*. The interaction between *complexity* and *condition;* $F(1, 71) = 4.934$, $p = .030$, $\eta_p^2 = .065$, indicated that *n*-back performance on disjunctive judgments took substantially longer than MP judgments in the control condition (*4,612ms-Disjunctives vs. 3,967ms-MP*) whereas the difference was notably smaller in the experimental condition (*3,368ms-Disjunctives vs. 3,136ms-MP*). The implications of these findings will be addressed in the discussion section. All remaining interactions were not significant (all $p > .1$).

Belief and Logic Instruction

A 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(Experimental/Control Condition) mixed design ANOVA was used for the second set of analyses measuring the effects under belief and logic instruction. The results produced a main effect of *instruction* (Belief/Logic); $F(1, 71) = 22.482$, $p < .001$, $\eta_p^2 = .240$, with better performance on logic based judgments than belief judgments *(92% logic vs. 88% belief)*. There was a main effect of *problem* type (Conflict/No-Conflict); $F(1, 71) = 37.058$, $p < .001$, $\eta_p^2 = .343$, with conflict items producing lower accuracy scores than no-conflict items *(87% vs. 93%)*. There was also main effect of *complexity* (Disjunctives/MP); $F(1, 71) = 37.944$, $p < .001$, $\eta_p^2 = .348$, with poorer performance on disjunctives than MP judgments *(88% vs. 92%)* but no main effect of *condition*; $F(2, 71) = .909$, $p = .344$, $\eta_p^2 = .013$, signifying no substantial difference in accuracy scores across conditions *(91% experimental vs. 89% control)*.

The significant interaction between *instruction* and *problem* type was present; $F(1, 71) = 16.034$, $p < .001$, $\eta_p^2 = .184$, replicating the ubiquitous finding that conflict has less impact on conclusion validity *(91% L-conflict vs. 93% L-no-conflict)* compared to the believability of a conclusion *(84% B-conflict vs. 94% B-no-conflict)* (see Table 4.8).

There was a significant interaction between *problem* type and *complexity*; $F(1, 71) = 10.255$, $p = .002$, $\eta_p^2 = .126$, with conflict having less of an impact on disjunctive judgments *(86% Conflict vs. 90% No-conflict)* than MP judgments *(88% Conflict vs. 96% No-conflict)*. There was also an interaction between *instruction* and *complexity*; $F(1, 71) = 27.773$, $p < .001$, $\eta_p^2 = .281$, with a bigger difference in performance between disjunctives and MP judgments under logic instruction *(88% L-Disjunctives vs. 95% L-MP)* than under belief instruction *(88% B-Disjunctives vs. 89% B-MP)*.

Finally, there was a 3 way interaction between *instruction, problem* type and *condition,* $F(1, 71) = 9.035$, $p = .004$, $\eta_p^2 = .113$, which follow up analyses confirmed was carried by the experimental condition producing a significant interaction between *instruction* and *problem* type; $F(1, 36) = 18.727$, $p < .001$, $\eta_p^2 = .342$, compared to no interaction for the control condition; $F(1, 35) = .745$, $p = .394$, $\eta_p^2 = .021$. The results suggest that in the experimental condition the effect of conflict is greater for belief judgments *(82% B-Conflict vs. 95% B-No-conflict)* than logic judgments *(94% L-Conflict vs. 94% L-No-conflict),* whereas in the control condition the effect of conflict is similar for belief (*85% B-Conflict vs. 91% B-No-conflict)* and logic judgments *(89% L-Conflict vs. 92% L-No-conflict)* (see Table 4.8).

Table 4.8

Average Accuracy and Latency scores for Belief and Logic Instruction on Conflict and No-conflict Problems, for MP and Disjunctive arguments. Results exclude below chance scores and include correct only latencies. Experiment 8: Experimental Condition (N = 37), Control condition (N=36).

Variable	Belief Conflict		Belief No-conflict		Logic Conflict		Logic No-conflict		Overall Means
	MP	Disj	MP	Disj	MP	Disj	MP	Disj	
Experimental									
Response Accuracy (%)	80	84	98	91	96	91	99	89	91
Latency (ms)	6,615	6,318	6,244	5,971	6,464	6,880	6,290	6,832	6,452
Control									
Response Accuracy (%)	84	86	93	90	92	85	95	89	89
Latency (ms)	6,446	6,336	5,717	6,090	5,993	6,371	5,358	6,366	6,085
Mean Accuracy (%) (across each cell)	82	85	96	91	94	88	97	89	
Mean Latency (ms) (across each cell)	6,531	6,327	5,981	6,031	6,229	6,626	5,824	6,599	

A mixed design ANOVA on response latencies produced no main effect of *instruction*; $F(1, 71) = .796$, $p = .375$, $\eta_p^2 = .011$, showing no significant difference in response latencies for belief *(6,218ms)* and logic *(6,320ms)* judgments, but there was a main effect of *problem* type*;* $F(1, 71) = 9.097$, $p = .004$, $\eta_p^2 = .114$, with conflict items taking longer than no-conflict items (*6,428ms vs. 6,109ms*). There was also a main effect of *complexity;* $F(1, 71) = 6.009$, $p =.017$, $\eta_p^2 = .078$, demonstrating quicker response times to MP arguments compared to disjunctive arguments (*6,141ms vs. 6,396ms*) but there was no main effect of *condition*; $F(2, 71) = 1.217$, $p = .274$, $\eta_p^2 = .017$, showing no significant difference in latencies across conditions. However a marginal interaction between *instruction* and *condition;* $F(1, 71) = 3.949$, $p = .051$, $\eta_p^2 = .053$, revealed that the difference between the control and experimental group for latencies was larger for logic judgments *(6,022ms control vs. 6,615ms experimental)* than for belief judgments

(6,147ms control vs. 6,287ms experimental). Analyses on the separate conditions confirmed that the main effect of *instruction* was only significant in the experimental condition; $F(1, 36) = 4.478$, $p = .041$, $\eta_p^2 = .111$, indicating that the experimental design slowed down performance on logic judgments more so than belief judgments (*6,615ms vs. 6,287ms*). The latencies also produced significant interactions between *problem* type and *complexity;* $F(1, 71) = 4.161$, $p = .045$, $\eta_p^2 = .055$, suggesting that conflict had less of an impact on disjunctive latencies *(6,477ms conflict vs. 6,315ms no-conflict)* than latencies for MP judgments *(6,380ms conflict vs. 5,903ms no-conflict).* Finally, an interaction between *instruction* and *complexity;* $F(1, 71) = 18.414$, $p < .001$, $\eta_p^2 = .206$, showed a bigger difference in response times between MP and disjunctive judgments under logic instruction *(6,027ms L-MP vs. 6,613ms L-Disjunctives)* than under belief instruction *(6,256ms B-MP vs. 6,179ms B-Disjunctives)* (see Appendix C for full data sets).

The next set of analyses reports the main effects and interactions between each experimental variable and the CRT measure. A 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(High and Low CRT) mixed design ANOVA carried out on accuracy scores revealed a main effect of CRT; $F(2, 71) = 4.914$, $p = .030$, $\eta_p^2 = .065$, with high CRT scorers producing better accuracy scores *(93%)* compared to the low CRT group *(87%).* However, there was no interaction between *instruction, problem* and *CRT*; $F(1, 71) = .019$, $p = .891$, $\eta_p^2 < .001$, or *problem* type, *complexity* and *CRT*; $F(1, 71) = .577$, $p = .450$, $\eta_p^2 = .008$, as reported in the previous experiment and there were no other significant interactions to note (all $p > .05$).

The Global Span (GS) scores only revealed a significant correlation with belief no-conflict items; $r(71) = .290$, $p = .013$, suggesting that the higher the GS score the better the performance on these items. Surprisingly the correlations with the SIE were all positive in this experiment; however none of which were at a reliably accepted level of significance.

Table 4.9
Correlations for Global Span scores and Stoop Interference Effect, measured against each Instructional set and for both Problem Types. Results exclude below chance accuracy scores.

GS = Global Span; SIE = Stroop Interference Effect

	Belief No-Conflict	Belief Conflict	Overall Belief	Logic No-Conflict	Logic Conflict	Overall Logic
GS	.290*	.046	.135	.212	.151	.187
SIE	.228	.171	.206	.210	.222	.228

Key: ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

4.3.3 Discussion

The principle objective of the current experiment was to examine the impact of *n*-back on performance relative to a control task that did not have any memory demands. Additionally, we wanted to evaluate whether the *n*-back task would have its greatest impact on logic judgments. The results comparing *n*-back instruction (experimental) with matching instruction (control) revealed that the experimental condition was sufficiently more demanding on performance than the control condition. The results also demonstrated that the complexity of the argument affected both the ability to recall and to match font characteristics, suggesting that both the *n*-back and the matching task interfered with structural processing. Moreover, embedded conflict within a problem marginally impacted matching and updating working memory, which proved significantly harder and took longer when the problems were more complex. Although accuracy performance was better in the control condition, participants took significantly longer to respond. One explanation is that with the *n*-back task participants are required to consistently keep an active representation of the last word in mind from trial to trial, whereas with the matching task there is no active representation, instead participants are required to refer back to the first premise in order to make a match with the last word in the concluding statement, and this takes longer to complete.

As specified earlier, our primary objective was to determine whether the *n*-back task was having more of an impact on logic judgments. Separate analysis on belief and logic instruction replicated all the main effects from Experiments 7, with belief judgments being more prone to errors, conflict problems harder to solve and disjunctive arguments producing poorer accuracy scores; as well as conflict having a larger influence on the believability of a conclusion. Follow up analyses revealed that the increased effect of conflict on belief judgments was only statistically significant in the Experimental condition, even though the means from the Control group tell a similar story, suggesting that logic impacts on belief judgments to a greater extent when demands on WM are increased.

The findings from the current experiment conflict with our initial hypothesis which proposed that the *n*-back task would specifically impact on logic judgments. Instead it appears that a task demanding of WM resources such as the *n*-back has a greater impact on the already more challenging belief based judgments. This however, leaves some questions unanswered, for example; in what way is the *n*-back task affecting the ability to resolve belief/logic conflict? And why are logic judgments taking longer to complete? We will consider these questions in more detail in the general discussion but in short we conjecture that there are two routes to a logical solution. One route is an independent Type 1 process completing first and creating an intuitive cue, perhaps accompanied by a feeling of rightness (Thompson, Prowse Turner & Pennycook, 2011), based on the logical structure of the argument. This output is what requires inhibiting and causes a Type 1/Type 2 conflict when instructed to reason on the basis of beliefs.

Inhibition of the intuitive logical cue is an effortful Type 2 process and the WM demands created by the *n*-back task affects the ability to keep the logic output inhibited.

The second route to a logical solution is a Type 2 process that runs parallel to the route required for a belief based response. When instructed to reason explicitly on the basis of logic the second route is interfered with by the *n*-back task which can explain why logical inferences take longer to complete. However, these logical inferences are still simple and are therefore less prone to error than belief based judgments. At this stage we infer that the *n*-back task effects both belief and logic based judgments but in different ways.

The purpose of this chapter, as discussed in the introduction, was to investigate the role of WM and the possibility that belief and logic based judgments pull on distinct executive resources. We predicted that the *n*-back task would have its greatest impact on logic and that inhibition (Experiment 9) will impact more on belief based judgments. Experiment 8 confirmed that the *n*-back task does impact on logic judgments, specifically the length of time it takes to produce a logical output, but it also demonstrated that increasing WM demands impacts on the ability to resolve conflict in favour of a believable response.

We infer that the WM demands created by the extra response alternative requiring participants to store, recall and update font characteristics were making it harder to inhibit the intuitive logical output, in support of the previous assumption that belief judgments pull on inhibition resources. In order to investigate whether inhibition plays this key role in the effect of conflict on belief based judgments, Experiment 9 directly examines the executive function with a modified version of the Stroop task (Stroop, 1935) integrated with the reasoning task.

4.4 Experiment 9

The following experiment will use a similar methodology to that previously described, expect this time an additional instructional condition is introduced to draw on the executive function of inhibition. The inhibition of pre-potent responses is a difficult resource to measure (see Friedman & Miyake, 2004, for discussion) but it is vital for understanding executive control and in this instance, understanding its role in reasoning under conflict. We conjectured that an efficient way of gauging the effect of inhibition on reasoning was to customise a version of the Stroop task by incorporating it into the reasoning task and measuring the subsequent impact on accuracy and response latencies. We predict that inhibition is the central resource required for dealing with conflict between a fast completing logical response and producing an effortful belief based output. Therefore, increasing the inhibitory demands of the task should have its greatest impact on belief judgments. We will also examine individual differences in the same way as Experiments 7 and 8.

The Stroop Task

The Stroop task is one designed to provoke conflicting responses. John Ridley Stroop first introduced the technique in 1935, which required participants to name the ink colour of a coloured word that was either congruent or incongruent to the actual word. The prepotent response is to say the colour that is written rather than the ink colour; therefore a correct response on incongruent trials involves inhibiting the automatic inclination to read the word presented. In Experiment 9 we manipulate the ink colour of the last word in each problem presented and ask participants to make judgments based on the validity or believability of a conclusion or on the ink colour of the last word in the conclusion (see Table 4.10 & 4.11). The integration of a Stroop condition is designed to increase inhibitory demands which will allow us to evaluate the subsequent effect this has on belief and logic based judgments.

4.4.1 Method

Participants

A total of 112 participants took part in exchange for two course credits. Ninety two females and 20 male participants were randomly assigned to either the Experimental condition (Stroop-Inhibition) or the Control condition (No inhibition).

Design

This experiment used a 3 x 2 x 2 x 2 mixed design which required participants to judge the validity or believability of the conclusion or the ink colour of the last word in the conclusion, creating three *instructional* sets. Participants were presented with both conflict and no-conflict problems (*problem type*) in the form of MP and disjunctive arguments (*complexity*) in either the Experimental or Control condition (*condition*) as the between subject factor.

Materials

Reasoning Task:

The same 192 conditional arguments from Experiment 7 & 8 were used, with the 64 *n*-back trials converted into Colour-Stroop questions. Each participant received 64 belief trials, 64 logic trials and 64 Colour-Stroop trials. The trials in each condition consisted of 32 MP and 32 Disjunctive arguments. For the belief and logic trials, 16 of each set were conflict problems, with valid/unbelievable or invalid/believable conclusions and the remaining 16 were no-conflict problems with conclusions that were valid/believable and invalid/unbelievable (see Appendix B for materials).

Experimental Manipulation – Colour-Stroop Task:

The Colour-Stroop questions were created by changing the ink colour of the last word in the conclusion. Primary and secondary colours were predominately used (see Figure 4.1), along with black, pink and white, since these were easier to identify than the blended tertiary colours and prevented replication of colour matches (i.e. blue on orange then orange on blue).

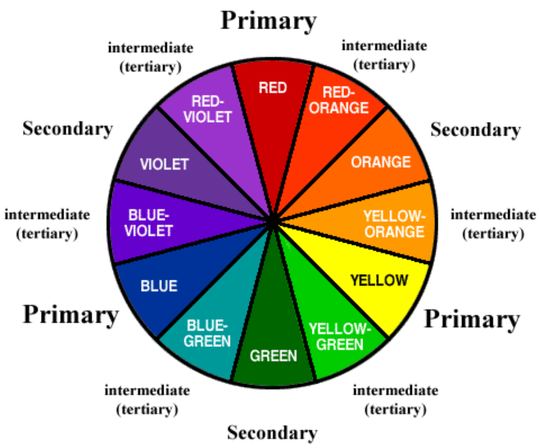


Figure 4.1
Colour Wheel showing the Primary, Secondary and Tertiary Colours.

In both the Experimental and Control conditions the participants were instructed to identify the ink colour of the last word, from the two response options given. In the experimental condition, the correct response always matched the ink colour of the final word in the conclusion and the alternative response option always matched the written word. For conflict trials this set up a prepotent response to choose the alternative, incorrect response option. Additionally, the correct response on conflict trials was either invalid/believable or valid/unbelievable. For no-conflict trials, the correct response always matched the ink colour as well as the believable/valid or unbelievable/invalid response, whilst the alternative response either matched the belief based or the logical response on half the trials and an unrelated colour in the remaining trials. For example, if the correct response was the colour `pink' which was also a valid/believable response then the alternative response would have been a colour that was the invalid/unbelievable response (on half the trials) or a colour that was unrelated to the believability or validity of the argument.

Table 4.10
Examples of the integrated Stroop trials for No-Conflict and Conflict Problems in the Experimental Condition.

No-conflict Problems		**Conflict Problems**	
If the fruit is a Strawberry, then it is red		If the fruit is a Strawberry, then it is purple	
Suppose fruit is a Strawberry		Suppose fruit is a Strawberry	
Does it follow that the fruit is red?		Does it follow that the fruit is red?	
s)green	k) **red**	s) **purple**	k) red
Either the Sky is blue or it is pink		Either the Sky is blue or it is pink	
Suppose the Sky is blue		Suppose the Sky is pink	
Does its follow that the Sky is not pink?		Does its follow that the Sky is blue?	
s) **pink**	k) blue	s) blue	k) **green**

*correct answer in bold.

For the control condition, the correct response on conflict trials matched the ink colour and the written word, whilst the alternative response was an unrelated colour that had no association with a belief or logic based response. The no-conflict trials were the same as those in the experimental condition (see Table 4.11).

Table 4.11

Examples of the integrated Stroop trials for No-Conflict and Conflict Problems in the Control Condition.

No-conflict Problems	**Conflict Problems**
If the fruit is a Strawberry, then it is red	If the fruit is a Strawberry, then it is purple
Suppose fruit is a Strawberry	Suppose fruit is a Strawberry
Does it follow that the fruit is red?	Does it follow that the fruit is red?
s)green k) **red**	s) yellow k) **red**
Either the Sky is blue or it is pink	Either the Sky is blue or it is pink
Suppose the Sky is blue	Suppose the Sky is pink
Does its follow that the Sky is not pink?	Does its follow that the Sky is blue?
s) **pink** k) blue	s) **blue** k) green

*correct answer in bold.

Measures of Individual Differences:

On completion of the Colour-Stroop reasoning task, the short **Stroop Task**, **Operation Span Task** and **CRT** were administered to each participant (see Experiment 7 for more details).

Procedure

Participants were randomly assigned to the Experimental or Control condition and tested in groups of four (see Experiment 7). They were asked to complete 12 practice trials before being presented with 192 problems in an unblocked design with the optional respite at the half way point. The experiment was presented on a computer screen with specific instructions for the colour-Stroop trials and the same instructions for the reasoning task as those presented in Experiments 7 and 8. The randomisation of the trials ensured that participants would remain unaware as to whether they would need to answer according to logic, beliefs or the colour of the last word, until the response options were revealed.

4.4.2 Results

A total of 16 participants were eliminated for performing below chance on conflict items, three from the Experimental condition (N = 50) and 13 from the Control condition (N = 46). Equivalent to the preceding two experiments, the first set of analyses presents the results under colour-Stroop instruction, once again to ensure that the experimental condition proved sufficiently more demanding than the control condition; followed by the independent analyses under belief and logic instruction (for full data sets see Appendix C).

Colour- Stroop Instruction

A 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x2(Experimental/Control Condition) mixed design ANOVA was carried out on Arcsine transformed accuracy data under colour-Stroop instruction. The *problem* type and *complexity* variables indicate the type of problem that accompanies the colour-Stroop trials. The results revealed a main effect of *problem* type (Conflict/No-Conflict); $F(1, 94) = 37.066$, $p < .001$, $\boldsymbol{\eta_p^2} = .283$, with Stroop trials accompanying conflict problems producing lower accuracy scores than those accompanying no-conflict problems *(96% vs. 99%)*. There was a significant main effect of *complexity* (Disjunctives/MP); $F(1, 94) = 34.873$, $p < .001$, $\boldsymbol{\eta_p^2} = .271$, where Stroop trials accompanying MP judgments produced lower scores than those accompanying disjunctives judgments *(97% vs. 99%)*. There was also a main effect of *condition;* $F(2, 94) = 5.845$, $p = .018$, $\boldsymbol{\eta_p^2} = .059$, which highlighted poorer performance in the experimental condition compared to the control condition *(97% vs. 98%).* Results also revealed a significant interaction between *problem* type and *complexity;* $F(1, 94) = 21.597$, $p < .001$, $\eta_p^2 = .189$, indicating a bigger difference in performance between MP and disjunctives for conflict items *(94% MP vs. 98% disjunctives)* compared to no-conflict items (*99% MP vs. 99% disjunctives).* There was an interaction between *problem* type and *condition*; $F(1, 94) = 54.942$, $p < .001$, $\eta_p^2 = .369$, which separate analyses revealed that the main effect of *problem* type was only present in the experimental condition*;* $F(1, 49) = 93.371$, $p < .001$, $\eta_p^2 = .656$, showing poorer performance on Stroop trials accompanying conflict problems compared to no-conflict problems *(94% vs. 100%).* There was also an interaction between *complexity* and *condition*; $F(1, 94) = 115.982$, $p < .001$, $\eta_p^2 = .552$, which further analyses showed that performance was better for Stroop trials associated with MP judgments *(99%)* in the control condition compared to those associated with disjunctives judgments *(97%)* whereas the reverse was true in the experimental condition *(99% disjunctives vs. 94% MP).* Furthermore, the 3 way interaction between *problem* type, *complexity* and *condition*; $F(1, 94) = 69.692$, $p < .001$, $\eta_p^2 = .426$, was driven by the significantly larger interaction between *problem* and *complexity* for the experimental condition; $F(1, 49) = 106.408$, $p < .001$, $\eta_p^2 = .685$, compared to the control condition; $F(1, 45) = 5.447$, $p = .024$, $\eta_p^2 = .108$. The results revealed that conflict had a bigger effect on Stroop trials accompanying MP judgments *(98% no-conflict vs. 89% conflict)* compared to those accompanying disjunctives judgments *(100% no-conflict vs. 99% conflict)* in the experimental condition. In the control condition there was no effect of conflict on disjunctive judgments *(97% no-conflict vs. 97% conflict)* but performance was slightly better on MP conflict problems (*99%)* compared to no-conflict problems *(98%),* under colour-Stroop instruction.

Table 4.12

Average Accuracy and Latency scores on Conflict and No-conflict Problems for MP and Disjunctives items under Colour-Stroop Instruction, across both Conditions. Results exclude below chance scores and include correct only latencies. Experiment 9: Experimental Condition (N = 50), Control condition (N=46).

Variable	Conflict		No-conflict		Overall means
	MP	Disjunctives	MP	Disjunctives	
Experimental					
Response Accuracy (%)	89	98	99	100	97
Latency (ms)	4,191	3,971	4,201	4,099	4,116
Control					
Response Accuracy (%)	99	97	98	97	98
Latency (ms)	2,893	2,801	3,323	3,146	3,041
Mean Accuracy (%) (across each cell)	94	98	99	99	
Mean Latency (ms) (across each cell)	3,542	3,386	3,762	3,623	

The latency data for colour-Stroop instruction produced a main effect of *problem* type*;* $F(1, 94) = 7.145$, $p = .009$, $\eta_p^2 = .071$, with Stroop trials accompanying no-conflict problems taking longer than conflict problems *(3,693ms vs. 3,464ms).* This could be due to the fact that no-conflict problems strongly cue a response that is compatible with both belief and logic and therefore takes longer to shift to a response based on colour. There was a main effect of *complexity;* $F(1, 94) = 5.400$, $p = .022$, $\eta_p^2 = .054$, showing that under colour-Stroop instruction arguments presented as MP took longer than those presented as disjunctives *(3,652ms vs. 3,504ms).* Finally there was a main effect of *condition*; $F(2, 94) = 27.137$, $p < .001$, $\eta_p^2 = .224$, indicating that overall, performance in the experimental condition was significantly slower than the control condition *(4,116ms* *vs. 3,041ms),* consistent with the experimental condition being more demanding on inhibition.

Belief and Logic Instruction

The accuracy data was analysed using a 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(Experimental condition/Control condition) mixed design ANOVA on Arcsine transformed data for belief and logic instruction. Results produced a main effect of *instruction* (Belief/Logic); $F(1, 94) = 22.637$, $p < .001$, $\boldsymbol{\eta_p^2} = .194$, with logic judgments generating higher accuracy scores compared to belief judgments *(93% vs. 89%)*. There was a main effect of *problem* type (Conflict/No-Conflict); $F(1, 94) = 56.043$, $p < .001$, $\boldsymbol{\eta_p^2} = .374$, with poorer performance on conflict items compared to no-conflict items *(88% vs. 94%)*. There was a main effect of *complexity* (Disjunctives/MP); $F(1, 94) = 50.773$, $p < .001$, $\boldsymbol{\eta_p^2} = .351$, where MP produced higher accuracy scores than disjunctive judgments *(94% vs. 88%)* but there was no main effect of *condition*; $F(2, 94) = .010$, $p = .919$, $\boldsymbol{\eta_p^2} < .001$, showing that performance between the experimental and control condition was not significantly different *(91% vs. 91%)*.

There was a significant interaction between *instruction* and *problem* type; $F(1, 94) = 8.467$, $p = .004$, $\boldsymbol{\eta_p^2} = .083$, revealing that the belief/logic conflict had more of an impact on judgments of the conclusions believability *(85% B-conflict vs. 94% B-no-conflict)* than its logical validity *(91% L-conflict vs. 95% L-no-conflict)*. There was also a 3 way interaction between *instruction, problem* type and *condition;* $F(2, 94) = 3.710$, $p = .057$, $\boldsymbol{\eta_p^2} = .038$, which separate analyses on both conditions revealed that the interaction between *instruction* and *problem* was present in the experimental condition; $F(1, 49) = 8.871$, $p = .004$, $\boldsymbol{\eta_p^2} = .153$, but absent in the control condition; $F(1, 45) = .789$, $p = .379$, $\boldsymbol{\eta_p^2} = .017$. This demonstrated that the effect of conflict was enhanced for belief instruction *(82% conflict vs. 94% no-conflict)* compared to logic instruction *(93% conflict vs. 94% no-conflict)* in the experimental condition. In the control condition the effect of conflict was comparable between belief instruction (*87% conflict vs. 94% no-conflict)* and logic instruction *(89% conflict vs. 95% no-conflict)*. A significant interaction between *problem* type and *complexity*; $F(1, 94) = 17.322$, $p < .001$, $\eta_p^2 = .156$, suggested that the impact of conflict was larger for MP judgments *(89% conflict vs. 97% no-conflict)* than disjunctives *(85% conflict vs. 90% no-conflict)* (see Table 4.13).

Table 4.13

Average Accuracy and Latency scores for Belief and Logic Instruction on Conflict and No-conflict Problems, for MP and Disjunctive arguments. Results exclude below chance scores and include correct only latencies. Experiment 9: Experimental Condition (N = 50), Control condition (N=46).

Variable	Belief Conflict		Belief No-conflict		Logic Conflict		Logic No-conflict		Overall Means
	MP	Disj	MP	Disj	MP	Disj	MP	Disj	
Experimental									
Response Accuracy (%)	82	82	97	90	95	90	97	91	91
Latency (ms)	6,018	6,126	5,375	5,917	6,473	6,628	5,426	5,778	5,968
Control									
Response Accuracy (%)	88	85	97	90	92	85	98	91	91
Latency (ms)	5,753	6,235	5,155	5,876	5,560	6,326	5,245	6,023	5,772
Mean Accuracy (%) (across each cell)	85	84	97	90	94	88	98	91	
Mean Latency (ms) (across each cell)	5,886	6,181	5,265	5,897	6,017	6,477	5,336	5901	

Finally, there was a three way interaction between *instruction, problem* and *complexity;* $F(1, 94) = 6.131$ $p = .015$, $\eta_p^2 = .061$. Separate analyses on belief and logic instruction revealed a significant interaction between *problem* type and *complexity* under belief instruction; $F(1, 94) = 19.695$ $p < .001$, $\eta_p^2 = .173$, demonstrating a bigger effect of conflict on MP judgments *(97% no-conflict vs. 85% conflict)* compared to disjunctive judgments *(90% no-conflict vs. 84% conflict)*. Whereas under logic instruction there was no interaction; $F(1, 94) = 1.451$ $p = .231$, $\eta_p^2 = .015$, showing that there was no differential effect of conflict on MP judgments *(98% no-conflict vs. 94% conflict)* compared to disjunctive judgments *(91% no-conflict vs. 88% conflict)*.

A mixed design ANOVA on response latencies, produced no main effect of *instruction*; $F(1, 94) = 2.150$, $p = .146$, $\eta_p^2 = .022$, revealing no significant difference in latencies between belief *(5,807ms)* and logic *(5,933ms)* instruction. There was, however, a main effect of *problem* type; $F(1, 94) = 34.661$ $p < .001$, $\eta_p^2 = .269$, with conflict items taking longer to complete than no-conflict items (*6,140ms vs. 5,600ms*) and a main effect of *complexity; F*(1, 94) = 38.171, $p < .001$, $\eta_p^2 = .289$, indicating faster response latencies to MP judgments compared to disjunctive judgments (*5,626ms vs. 6,114ms*). The interaction between *complexity* and *condition; F*(1, 94) = 6.329, $p = .014$, $\eta_p^2 = .063$, suggested that there was bigger effect of condition on MP latencies (*5,428ms control vs. 5,823ms experimental)*, compared to the latencies on disjunctive judgments (*6,114ms control vs. 6,112ms experimental)*. There was, however, no main effect of *condition*; $F(2, 94) = .359$, $p = .550$, $\eta_p^2 = .004$, *(5,772ms control vs. 5,968ms experimental)*.

Lastly, there was a 3 way interaction between *instruction*, *problem* type and *condition;* $F(1, 94) = 7.290$, $p = .008$, $\eta_p^2 = .072$. Analysis on each condition revealed that the *instruction* by *problem* type interaction was present in the experimental condition; $F(1, 49) = 9.586$, $p = .003$, $\eta_p^2 = .164$, showing that conflict had its greatest impact under logic instruction *(6,551ms-L-conflict vs. 5,602ms-L-no-conflict)* compared to belief instruction (*6,072ms-B-conflict vs. 5,646ms-B-no-conflict)*. However, the interaction was absent in the control condition; $F(1, 45) = .763$, $p = .387$, $\eta_p^2 = .017$, suggesting that there was no differential effect of conflict on logic instruction *(5,943ms-L-conflict vs.*

5,634ms-L-no-conflict) compared to belief instruction (*5,994ms-B-conflict vs. 5,516ms-B-no-conflict)* in the control condition. In other words, logic-conflict items took longer to complete when participants were also required to engage in the colour-Stroop task.

Accuracy scores on the CRT were collapsed across both the experimental and control conditions and only significant main effects and interactions between experimental variables and the CRT are reported. A 2(Belief/Logic) x 2(Conflict/No-Conflict) x 2(MP/Disjunctives) x 2(High and Low CRT) mixed design ANOVA was carried out on accuracy scores which produced a main effect of CRT; $F(2, 94) = 8.191$, $p = .005$, $\eta_p^2 = .080$, with higher accuracy scores from the high CRT group *(94%)* compared to the low CRT group *(88%)*. There was a significant interaction between *instruction* and *CRT*; $F(1, 94) = 5.577$, $p = .020$, $\eta_p^2 = .056$, with high CRT scorers performing equally well on belief and logic judgments *(93% vs. 94%)* compared to low CRT scorers, who performed less well on belief judgments compared to logic judgments *(85% vs. 91%)*. There was a significant interaction between *problem* type and *CRT*; $F(1, 94) = 4.753$, $p = .032$, $\eta_p^2 = .048$; which showed that the low group performed less well on conflict items compared to no-conflict items *(83% vs. 93%)* whilst the high group produced similar scores for both problem types *(92% conflict vs. 95% no-conflict)*. Finally, there was a marginal 3 way interaction between *instruction*, *problem* type and *CRT*; $F(1, 94) = 3.670$, $p = .058$, $\eta_p^2 = .038$, which separate analysis carried out on each group revealed a significant interaction between *instruction* and *problem* type for the low CRT group; $F(1, 50) = 10.912$, $p = .002$, $\eta_p^2 = .179$, demonstrating that conflict had a greater impact on belief judgments *(78% B-conflict vs. 92% B-no-conflict)* compared to logic judgments *(89% L-conflict vs. 93% L-no-conflict)*. Whereas there was no significant interaction between *instruction* and *problem* type for the high CRT group; $F(1, 44) = .493$, $p = .486$, $\eta_p^2 = .011$; showing that the effect of conflict was comparable for belief judgments *(91% B-conflict vs. 95% B-no-conflict)* and logic judgments *(93% L-conflict vs. 96% L-no-conflict)*. The latency data produced no main effect or interactions with CRT (all p > .05).

Table 4.14
Average Accuracy scores for Conflict and No-conflict Problems, under both Instructions across High and Low CRT. Results exclude below chance accuracy scores. High CRT (N = 45), Low CRT (N=51).

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No-conflict	Conflict	No-conflict	
High CRT Average Response (%)	91	95	93	96	94
Low CRT Average Response (%)	78	92	89	93	88
Mean Accuracy (%) (across each cell)	85	94	91	95	

As in experiments 7 & 8, a correlational analysis was carried out on the data from both the Operation Span Task and the short Stroop Task (see Table 4.15). Results revealed a significant positive correlation between Global Span (GS) scores and belief items; $r(94) = .268$, $p = .008$, specifically belief-conflict items; $r(94) = .256$, $p = .012$, showing that the higher the GS score the better the performance on these problems, which extended to logic no-conflict items; $r(94) = .212$, $p = .038$.

The data from the short Stroop task produced a significant negative correlation between the SIE and overall performance on belief trials; $r(94) = -.244$, $p = .016$ but specifically with belief conflict items; $r(94) = -.288$, $p = .004$, suggesting that those who experience less Stroop interference perform better on belief problems.

Table 4.15
Correlations for Global Span scores and Stoop Interference Effect, measured against each Instructional set and for both Problem Types. Results exclude below chance accuracy scores.

GS = Global Span; SIE = Stroop Interference Effect

	Belief No-Conflict	Belief Conflict	Overall Belief	Logic No-Conflict	Logic Conflict	Overall Logic
GS	.178	.256*	.268**	.212*	.028	.115
SIE	-.009	-.288**	-.244*	.033	-.185	-.103

Key: ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

4.4.3 Discussion

The aim of Experiment 9 was to examine the role of inhibition when reasoning under belief and logic instruction. More specifically we were interested in determining whether increasing the inhibitory demands of the task, by incorporating a colour-Stroop instruction manipulation, would help evaluate its role in the effect of conflict on belief based judgments.

The effects from the previous experiments in this chapter were replicated, confirming that belief judgments are more prone to errors, conflict problems are harder to solve and MP judgments produce better accuracy scores, along with the prevalent finding that conflict between belief and logic impacts more on the believability of the conclusion for both accuracy and latency scores (Handley et al., 2011). One of the key results from the present study was that the Experimental condition increased this effect of conflict under belief instruction suggesting that logic impacts on belief judgments to a greater extent when inhibitory demands are increased by the Stroop Task.

Separate analysis on colour-Stroop instruction revealed a main effect of condition, confirming that the experimental design was more taxing than the control condition. Furthermore, the results demonstrated that under colour-Stroop instruction MP judgments were harder and look longer to complete, and conflict had its greatest impact on these simpler judgments. This is contrary to experiments 7 & 8 that showed *n*-back instruction having a bigger effect on the more complex disjunctive arguments. One explanation for this relates to the reported automaticity of MP judgments (Reverberi, Burigo, & Cherubini, 2009) and the intuitive logical solution to these simple inferences. Let us imagine that the presentation of an `if p then q' premise triggers a logical response that completes quickly. This response would require inhibition until response options were displayed and further inhibition if instructed to respond based on the ink colour of the word, which was in conflict with colour triggered by the logical response (e.g. if the fruit is a strawberry then it is purple – purple would need inhibiting, see Figure 4.2). A task that pulls on the inhibition resource, such as a Stroop task, will impact on the ability to supress the automatically cued response increasing Stroop interference for MP judgments.

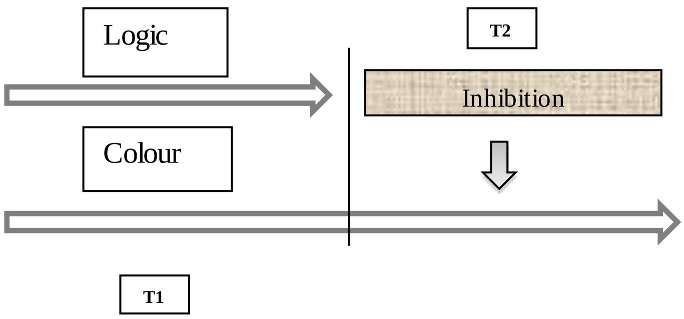


Figure 4.2
Illustrates the effect of Inhibition under colour-Stroop Instruction for MP arguments.

With disjunctive arguments perhaps the effect of conflict does not arise as early, consequently fewer resources are required to inhibit the logic response and therefore conflict has less of an impact on colour-Stroop judgments. In the general discussion we will consider in more detail the *problem* by *complexity* interaction displayed throughout this chapter.

The findings up to this point suggest belief judgments are effortful and tax on WM resources of inhibition. However, the experimental design appears to have a more considerable impact on the length of time it takes to complete logic-conflict items. As discussed in Experiment 8 we assume there are two routes to a logical output, one is intuitive and one requires Type 2 processing when given explicit instructions to reason logically. In this experiment the second route may be impacted on by the increased inhibitory demands created by the colour-Stroop task which interferes with the length of time it takes for a logical output to complete. This account will be discussed further in the general discussion.

As in Experiment 7, the CRT data revealed that logic interfered more with belief based judgments in the low CRT group. This fits with the argument that belief-conflict items require inhibitory control over the fast completing logical response. Those participants who are less reflective over a response available early (the logical output), will find belief-conflict items more challenging to complete.

Finally, results from the Operation Span task and short Stroop Task produced similar findings to Experiment 7; supporting the view that belief-conflict processing is effortful and correlates with measures of inhibition and WM capacity.

4.5 General Discussion

In the previous chapter we confirmed that both belief and logic based judgments require effortful processing, yet the greater impact of conflict on belief judgments and the larger effect of Random Number Generation (RNG) on logic judgments suggested that both pull on distinct executive resources. We conjectured that logic based responses completed quickly and required inhibiting in order to produce a response based on the believability of an argument. Logical judgments, on the other hand, seemed to draw on alternative Working Memory (WM) resources. The aim of this chapter was to investigate the role of WM in belief and logic based reasoning with the use of a memory *updating* task and a task that required the executive resources of *inhibition*. By employing a novel methodology and various individual differences measures we tested the function of these executive resources under distinct instructional sets. We hypothesised that the Stroop task would have a bigger impact on belief based judgments whilst the logical structure of the argument would be most affected by the *n*-back task.

The findings revealed that the effect of conflict was greater for belief judgments in the experimental condition for both Experiments 8 and 9, suggesting that the presence of both the *n*-back and colour-Stroop task increases the impact of logic on belief based judgments. However, responses under logic instruction took longer (Experiment 7) and the experimental design slowed down performance on logic judgments (Experiment 8), specifically logic-conflict items (Experiment 9) more so than those under belief instruction.

How then do we reconcile the findings that logic interferes with belief based judgments, yet a logical output takes longer to complete and is more impacted on by a task such as random number generation (Experiment 5)?

As proposed earlier we offer the possibility of two routes to a logical solution; the intuitive route and the deliberative route. The intuitive route is a Type 1 process which completes outside explicit awareness and is sensitive to the structure of a problem. This automatic logical output is probably accompanied by a Feeling of Rightness (Thompson, Prowse Turner & Pennycook, 2011) based on conceptual fluency (Morsanyi and Handley, 2012) which provides the reasoner with an intuitive logical cue. However, this cue is not enough to give a logical response, when instructed to reason logically, the second deliberative route must be engaged, which involves effortful Type 2 processing triggered in parallel to a belief based response.

We conjecture, therefore that the *instruction* by *problem* interaction continually displayed throughout this thesis is the result of a Type1/Types 2 conflict and the requirement to *inhibit* an intuitive logical output when instructed to reason according to beliefs. Inhibition of the logical output is a challenging task that taxes WM. Extra demands on WM created by additional response alternatives that require continuous updating (*n*-back) or inhibiting (Stroop), will impact on the ability to keep an intuitive logical solution inhibited (see Figure 4.3).

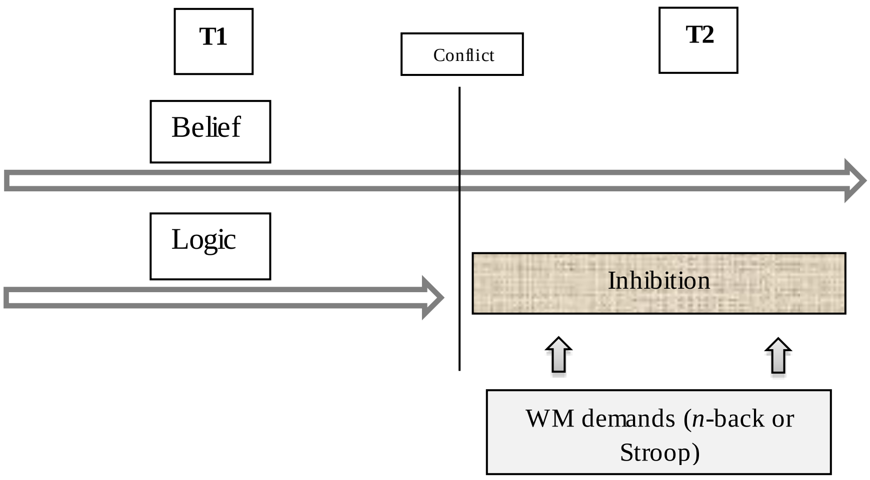


Figure 4.3
Illustrates the Default Interventionist model for ***Belief based responding*** *under* ***Belief instruction****. A logical output completes automatically and requires inhibiting in order for a Belief based output to complete. The n-back or Stroop task demands on WM which impacts on the ability to inhibit the Logical output.*

When instructed to reason logically, it takes explicit Type 2 reasoning or the `deliberate route' to a logical solution. The explicit demands of the *n*-back or Stroop task may interfere with the integration of premise information and model construction, which slows the response process but has less effect on accuracy scores due to the relatively simple nature of the problems (see Figure 4.4).

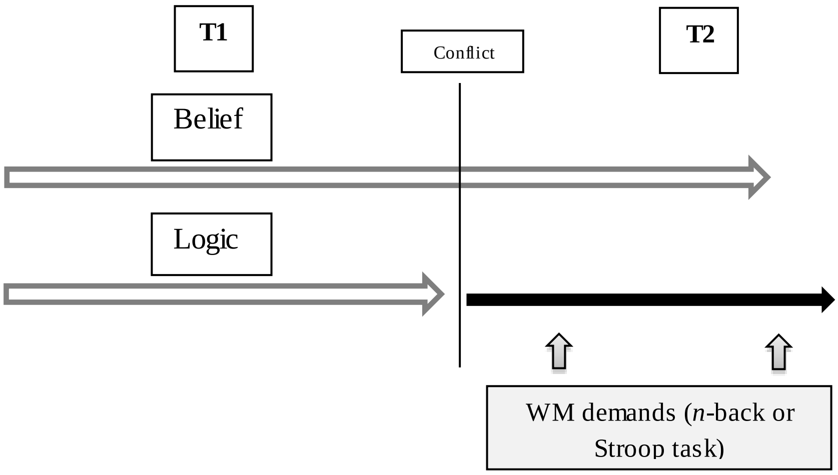


Figure 4.4
Illustrates the parallel processing involved when provided with ***Logic instructions****. In order to deliver an explicit Logical output, Type 2 processing is engaged. The demands of the n-back or Stroop task interrupts premise integration and model construction which affects the length of time it takes to produce output.*

Another consistent finding throughout this chapter is that conflict has less of an impact on disjunctive arguments for accuracy scores (also see Experiment 4, Chapter 3) and latency scores (Experiment 8). We posit that this is due to the relatively automatic nature of the MP inference, in support of the intuitive route. For example, if the presentation of an MP argument triggers an intuitive logical cue that requires inhibiting in order to produce a belief based response then the impact of conflict is likely to be greater for MP arguments, specifically under belief instruction as displayed in Experiment 9. This could also explain why MP arguments are more affected by colour-Stroop instruction, if for example the logical cue creates a Type1/Type 1 conflict with a colour cue, interfering with the ability to inhibit the ink colour for the written word (see Figure 4.2).

With the more complex disjunctive arguments we suppose that responses under belief and logic instruction both require some effortful Type 2 processing and thus conflict between the two occurs much later. With conflict occurring later, less inhibition of the logical response is required resulting in reduced interference with a belief based output.

The automaticity of the MP argument can assist in the interpretation of the CRT findings in this chapter. Both Experiment 7 and 8 showed that the impact of conflict was greater on beliefs for low CRT scorers and this effect was bigger with simple MP arguments (Experiment 7). This could suggest that those people more inclined to give an intuitive response over one that requires deliberation, will find the conflict between belief and logic more difficult when asked to respond on the basis of belief and more so when making judgments with more automatic MP inferences. This interpretation supports the idea that an intuitive logical cue impacts on belief based reasoning. Furthermore, in Experiments 7 and 9 the short Stroop task revealed that those who experience less Stroop Interference perform better on belief conflict items, indicating that *inhibition* is a key component in belief-conflict judgments. Nevertheless the fact that WM also correlated with belief-conflict reasoning, offers the likelihood that inhibition is not the only executive resource required for these types of judgments.

At this point, the results from the ID measures and the impact of the *n*-back on belief-conflict judgments (Experiment 8) encourages the move away from the idea that one type resource is required for one type of reasoning as proposed at the beginning of this chapter. For example, in Experiment 4 (Chapter 3) the CRT findings are not consistent with those in this current chapter. The previous findings showed that CRT was related to logical reasoning which we proposed was related to a person's ability to extract the underlying structure of a problem to arrive at the correct conclusion. Perhaps the ability to resist a rapidly available response in favour of a more considered response involves many processes which include both inhibition and the extraction of problem structure.

As mentioned earlier we move away from the idea that belief judgments pull on one executive resource and logic pulls on another, however, the behavioural data and ID measures in this chapter do point to inhibition as a key factor in belief-conflict reasoning. That is not to say that logic judgments are not effortful and demand WM resources (and possibly inhibition at times), we conjecture that when instructed to reason logically the explicit Type 2 processes involve the integration of premise information and model construction as explanation for longer latencies in Experiments 7 and the impact of RNG in Experiment 5 (Chapter 3). In the closing chapter we will discuss the two routes to a logical solution as an explanation for the findings in this thesis and discuss the implications of these findings in relation to the Dual Process theories on reasoning.

Chapter 5: General Discussion

5.1 Introduction

The main objective of this thesis was to test the *Default Interventionist* (DI) Dual Process account of belief-bias in human reasoning, using the novel methodology introduced by Handley, Newstead and Trippas (2011) which involved the manipulation of instruction. Handley et al. found that belief judgments were slower than logic judgments and the believability of a conclusion was impacted on by the logical structure of the argument (a finding that was reliably replicated in six of the experiments presented in this thesis, see section 5.2.1). They concluded that judgments based on the logical structure of the argument can be accomplished relatively rapidly and those made about beliefs require effortful processing which is inconsistent with a DI account of reasoning.

In this thesis we tested the interpretation that belief judgments are effortful and logic judgments can be accomplished quickly; using a number of manipulations designed to increase cognitive load when reasoning under belief and logic instruction. First we examined the effect of different types of secondary load on accuracy and latency performance of response with simple conditional arguments (Chapter 2). In Chapter 3 more complex arguments and various individual differences measures were introduced for examination. With the introduction of a between subjects manipulation, performance on the primary task was measured against the secondary task to see if there were any trade-offs between the two. Finally, in chapter 4 additional instructional conditions were incorporated in order to look at the role of WM and inhibition in belief and logic based reasoning. The experimental work supports a particular dual process interpretation which is discussed in section 5.4.

In this final chapter we start with a summary of the accuracy and latency data as a function of instruction (belief/logic), problem type (conflict/no-conflict) and complexity (MP/Disjunctives), in order to reveal any comparable findings with previous works. Next we look at the impact of secondary load across experiments, followed by the examination of distinct executive functions under each instructional set and the interpretation of the individual differences (ID) data. The theoretical implications of the findings are then discussed followed by our interpretation of the data in relation to a Parallel Competitive (PC) model. Support for the PC interpretation is then examined, followed by points for consideration in respect to research on biases. In the closing section, some potential areas of research development are discussed.

5.2 Summary of Experimental Findings

This experimental summary addresses each area of findings separately, starting with the accuracy and latency data summarised across instruction, problems type and complexity, in order to determine whether the pattern of results replicate earlier work by Handley et al. The next section looks at the critical set of findings regarding the impact of secondary load on reasoning performance followed by a section on the effect of manipulating executive demands. The final section explores individual differences and reasoning performance under each instructional set.

5.2.1 Accuracy and Latencies

Table 5.1

A summary table illustrating the presence of a significant main effect (Instruction) or interaction for both Accuracy and Latency data in each experiment. For example; in Experiment 1, there was an interaction between Instruction and Problem Type for Accuracy data but not for Latencies.

Experiment No.	Instruction		Problem		Complexity		Instruction * Problem		Instruction * Complexity		Instruction * Problem * Complexity		Complexity * Problem	
1	$A_{(L)}$	$L_{(B)}$	$A_{(NC)}$	$L_{(C)}$	N/A	N/A	A	-	N/A	N/A	N/A	N/A	N/A	N/A
2	-	$L_{(B)}$	$A_{(NC)}$	$L_{(C)}$	N/A	N/A	-	L	N/A	N/A	N/A	N/A	N/A	N/A
3	$A_{(L)}$	-	$A_{(NC)}$	$L_{(C)}$	N/A	N/A	-	-	N/A	N/A	N/A	N/A	N/A	N/A
4	$A_{(L)}$	-	$A_{(NC)}$	-	$A_{(MP)}$	$L_{(D)}$	A	-	-	-	-	-	A	-
5	$A_{(L)}$	$L_{(L)}$	$A_{(NC)}$	-	$A_{(MP)}$	$L_{(D)}$	A	-	-	-	-	-	-	-
7	$A_{(L)}$	$L_{(L)}$	$A_{(NC)}$	$L_{(C)}$	$A_{(MP)}$	-	A	-	A	-	-	-	A	-
8	$A_{(L)}$	-	$A_{(NC)}$	$L_{(C)}$	$A_{(MP)}$	$L_{(D)}$	A	-	A	L	-	-	A	L
9	$A_{(L)}$	-	$A_{(NC)}$	$L_{(C)}$	$A_{(MP)}$	$L_{(D)}$	A	-	-	-	A	-	A	-

Key: A = accuracy; L = latencies; $A_{(L)}$=Logic higher scores; $A_{(NC)}$= No-conflict higher scores; $A_{(MP)}$= MP higher scores; $L_{(B)}$=Belief take longer; $L_{(L)}$=Logic take longer; $L_{(C)}$=Conflict take longer; $L_{(D)}$=Disjunctives take longer.

From Table 5.1 what emerges is that in seven out of eight experiments there was an effect of instructional condition on accuracy scores. Belief judgments were consistently more difficult and resulted in more errors, consistent with the idea that belief based inferences are effortful. The table also shows that conflict problems were consistently harder and in six out of eight of the experiments they also took longer to complete. The interaction between these two factors (*instruction* and *problem*) revealed that conflict had a larger impact on belief judgments than logic judgments in six out of eight of the studies; for simple (Experiment 1) and complex problems (Experiment 4 – 9), in both a within-subjects and a blocked design (Experiment 5). Collectively these findings provide a clear replication of Handley et al.'s work. However, in terms of the latency data it is important to note that, although belief judgments took longer in Experiments 1 and 2, there was no difference between latencies for belief and logic judgments in Experiments 3 and 4 and belief judgments took less time than logic judgments in Experiments 5 and 7. The longer latencies for logic judgments are discussed in section 5. 4.

One possible explanation for the greater effect of conflict on belief judgments is that MP inferences can be drawn automatically (Reverberi, Burigo & Cherubini, 2009; Reverberi, Pischedda, Burigo & Cherubini, 2012). In Experiments 4 to 9 we included disjunctive arguments to see if the impact of conflict would extend to these more complex judgments and found that the impact was smaller for disjunctives compared to MP judgments. This finding was accompanied by a main effect of complexity showing that disjunctive judgments took longer to complete than MP judgments in four out of five studies. We conjecture that complexity has an impact on the extent to which belief judgments are influenced by a competing logical response; the more complex the inference the less opportunity for logic to interfere with beliefs (refer to Figure 5.4 section 5.5.1).

5.2.2 Secondary Load

One of the key objectives of the research presented in this thesis was to evaluate the impact of increasing executive demands on reasoning under different instruction. Table 5.2 provides a summary across all experiments in terms of the main effect of load/condition and its interaction with instruction, problem type and complexity. The table also includes Experiments 7 – 9 where a secondary task was not employed but additional instructional conditions were designed to increase executive demands; these experiments are discussed further in section 5.2.3.

Table 5.2

A summary table illustrating the presence of a significant main effect of Condition/Load and the interactions with Condition/Load for both Accuracy and Latency data in each experiment.

Experiment No.	Load/Condition		Instruction * Condition		Instruction * Problem * Condition		Complexity * Condition	
1	-	-	-	-	-	-	N/A	N/A
2	A	-	$A_{(L-m)}$	-	-	-	N/A	N/A
3	$A_{(m)}$	L	-	-	-	-	N/A	N/A
4	A	L	-	-	-	-	-	-
5	A	L	$A_{(L)}$	-	-	-	-	-
7	-	-	$A_{(B)}$	-	-	-	-	-
8	-	-	-	$L_{(L-m)}$	A	-	-	-
9	-	-	-	-	A	L	-	L

Key: A = accuracy; L = latencies; (m)=marginal; $A_{(L)}$=impact on logic accuracy; $A_{(B)}$= impact on belief accuracy; $L_{(L)}$=impact on logic latencies.

De Neys' (2006) Dot Memory task was used as the secondary load in Experiment 1, however this particular spatial WM task failed to impact on reasoning performance which we conjectured was because it did not tax the verbal-working memory component required in reasoning with conditional arguments (Handley, Capon, Copp & Harper, 2002). Therefore, Experiments 2 – 5 employed Random Number Generation (RNG) as the secondary load which has been shown to impact heavily on executive resources (Miyake et al., 2000).

The effect of RNG on performance in each instructional condition varied across experiments as seen in Table 5.2. Experiment 2 and 5 showed that RNG increased the difficulty of making validity judgments whilst Experiments 3 and 4 showed RNG as having a comparable effect on belief and logic judgments.

To explain the load effect in Experiment 2 we hypothesised that the requirement to remember the major premise after it disappeared, in order make a validity judgment, was increasing WM load. To control for this, Experiments 3 and 4 presented the major premise for three seconds before displaying the categorical premise and conclusion, keeping the full problem on the screen until a response was given. This resulted in a main effect of load which suggested that both belief and logic based judgments require effortful processing that demand WM resources.

Experiments 2 to 4 used a within-subjects design where belief and logic trials were randomly interspersed and participants were unaware of whether they would be responding according to beliefs or logic until the response options were presented. There are two possible consequences of this design; first is that it increases executive demands as a result of switching between instructional sets from one trial to the next. The second is that perhaps people reason logically through an argument before response options are presented because they are unaware of what the response requirements will be. Therefore, any effect of conflict is the result of these explicitly drawn inferences impacting on belief based judgments. To control for these possibilities, Experiment 5 used a blocked design which limited any switching demands by presenting each instructional set in blocks. The design also meant that participants were aware from the onset of each trial whether they were required to respond on the basis of belief or logic. The findings showed that the impact of conflict on belief judgments remained suggesting that participants continued to process the logical inference even when there was no requirement to do so. The blocked design also permitted the assessment of the random numbers generated as a function of the primary task. The results indicated no differential allocation of resources to the secondary task as a function of instruction, but in contrast to Experiment 4 load had its greatest impact on logical inferences. In other words; generating random numbers had its biggest effect on structural processing but the logical structure of an argument continued to interfere with the ability to reason on the basis of beliefs (see Table 5.1), even in a blocked design.

The evidence demonstrating an impact of RNG on validity judgments, coupled with the higher level of difficulty displayed with belief judgments suggests that both draw upon some aspect of effortful processing. The question is why does RNG only impact on logic in a blocked design yet belief judgments are impacted on when trials are interspersed? One potential explanation is that RNG mainly interferes with processing the structure of an argument, whilst switching between belief and logic judgments places additional demands on executive resources that effects belief based inferences. Through the removal of these switching demands RNG only interferes with judgments that require structural processing. Therefore, belief and logic judgments may both require effortful processing but pull on distinct executive resources.

5.2.3 Executive Functions

The differential effect of load displayed in Chapter 3, could be explained in terms of each instructional set demanding distinct executive resources. For example, logic based judgments may draw upon Working Memory (WM) to process the structure of an argument; whilst belief based judgments are impacted on by the requirement to inhibit one instructional set for another (switching demands). Consequently, in Chapter 4 the executive functions of inhibition and WM were examined in relation to reasoning performance. In Experiments 7 – 9 additional instructional sets were incorporated into the reasoning task in order to increase WM and inhibition demands. Experiments 7 and 8 used an original variant of the *n*-back task to examine memory updating and Experiment 9 used a Stroop task to assess the function of inhibition.

The results from Experiment 8 revealed that a task demanding of WM resources such as the *n*-back task has a bigger effect on the already challenging belief based judgments.

Additionally, when inhibitory demands were increased with the Stroop task in Experiment 9, the impact of logic on belief based judgments was also increased. Taken together these findings suggest, contrary to Experiment 5, that belief judgments are more demanding of executive resources. However, the latency data produced findings that were seemingly inconsistent with this interpretation. For example, in Experiment 7 logic judgments took longer than belief based judgments, in Experiment 8 the experimental condition had a bigger impact on logic latencies and Experiment 9 revealed that conflict had its biggest impact on logic latencies in the experimental condition. The accuracy data supports the idea that belief judgments require effortful processing whilst the latency data suggests that increased executive demands of the task impacts on the time it takes to reason explicitly on the basis of logic. The account presented in section 5.4 attempts to reconcile the accuracy and latency data.

5.2.4 Individual Differences

Research has shown that limitations to WM, intelligence and cognitive style can affect an individual's ability to resist biases (Macpherson & Stanovich, 2007; Stanovich, 2005; Stanovich and West, 2008; Stanovich & West, 1997). Stanovich and West (2008) found that individuals with high SAT scores displayed less belief-bias in syllogistic reasoning. They suggested that the ability for sustained inhibition of a heuristic response and the ability to separate imaginary situations from real world representations (see Stanovich & Toplak, 2012 for "cognitive decoupling") is a fundamental component in overcoming biased behaviour.

Experiment 4 examined whether individual differences in cognitive ability (as measured by the AH4; Heim, 1970) and cognitive style (as measured by the AOT; Macpherson & Stanovich, 2007 and the CRT; Frederick, 2005) were related to judgments under different instructions. Experiments 7, 8 and 9 examined more specific measures of inhibition and WM with the use of a short Stroop task and an Operation Span task.

In Experiment 4 both the AH4 and CRT were examined in the extent to which high and low scorers performed under belief and logic instruction. Across both measures there was a small but significant difference between the high and low groups under logical instruction. In contrast, Experiments 7 and 9 showed that high CRT scorers performed better than low CRT scorers on belief judgments when belief and logic conflicted. In addition the Global Span (GS) scores and Stroop Interference effect (SIE) correlated significantly with belief conflict items across both experiments whereas GS and SIE did not significantly correlate with performance under logic instruction.

Although Experiment 4 showed a stronger relationship with logic judgments, the effect size was small; whereas the evidence revealing belief judgments to be more resource demanding covered a broader range of measures including cognitive style, inhibition and WM. However, it is important to be cautious with interpretations of the ID measures, since the experiments presented in this thesis were not typical individual differences studies and the data was collapsed across experimental conditions. Furthermore, logic judgments were closer to ceiling level, thus reducing the likelihood of observing significant relationships with the ID measures.

The findings presented in this summary section have shown that belief based judgments are consistently more prone to error (see Table 5.1) and in six out of eight experiments are impacted on by the validity of the argument. Furthermore, whilst RNG impacts on the structural processes required for logical inferences (Experiment 5), switching between instructional sets may also increase the executive demands that effect belief based reasoning. The results from Experiments 7 – 9 indicate that enhancing the executive demands of the task, by incorporating extra response alternatives, increases the impact of logic on belief based judgments (Experiment 8 and 9). One interpretation for this impact is attributed to the demands of successful inhibition of a fast completing logical response in order to produce a belief based output. Inhibition of the logical output requires effortful Type 2 processes and these processes are impacted on by resource demanding tasks as shown in Experiments 8 and 9. However the longer latencies for logic judgments as seen in Experiment 5, 7, 8 and 9 suggests that when reasoning explicitly under logic instruction the extra demands of the task impact on the time it takes for a logical response to complete. The next two sections cover the theoretical implications of these findings in relation to Dual Processing accounts of reasoning, followed by an account of the data interpreted through an alternative version of the PC model.

5.3 Theoretical Implications

5.3.1 Default Interventionism

In the opening chapter of this thesis the DI dual process account of reasoning was introduced (Evans, 2008). This account posits that judgments based on prior knowledge or stereotypical information depends on intuitive Type 1 processes, which require inhibiting in order to produce the `correct' logical output, through the engagement of Type 2 processes. The override of these intuitive Type 1 responses will depend on instruction (Evans, Newstead, Allen, & Pollard, 1994) and WM or cognitive capacity (Stanovich & Toplak, 2012; Stanovich and West, 2008).

The behavioural data presented throughout this thesis does not fit well with the DI account; given that belief based judgments resulted in more errors and were more impacted on by the validity of an argument. Furthermore, the impact of conflict on belief judgments increased when the executive demands of the task increased and the ID measures from Experiment 7 and 9 correlated more strongly with belief-conflict trials. Taken together the data emphasises the role of Type 2 processing required for belief based judgments. Although there is extensive research that supports a DI account (see Chapter 1), these findings do not support the idea that belief based judgments are intuitive in nature. In addition, as discussed in Chapter 1, the DI account also fails to explain how conflict is detected early and this is a major limitation of the account.

Research on conflict detection has shown that reasoners are *implicitly* aware of two sources of information and can detect conflict between these sources; namely belief and logic based information for syllogistic problems (De Neys & Franssens, 2007; De Neys, Vartanian & Goel, 2008) where conflict leads to longer inspection times of these problems (Ball, Philips, Wade & Quayle, 2006). Conflict detection is also evident with base rate problems (De Neys & Glumicic, 2008) where the base rate information presented in a problem is unaffected by cognitive load (Franssens & De Neys, 2009). The question the DI account fails to resolve is; how can conflict be detected if logical and belief based responses rely on different `Types' of processes? An alternative model is one that implies the parallel processing of two responses as an appropriate explanation for early conflict detection.

5.3.2 Parallel Processing

The original Parallel Competitive (PC) model as taken from Sloman (1996) suggests that both Type 1 (associative) and Type 2 (rule based) processes are activated simultaneously which addresses the issue of conflict detection. However the way this particular PC model is presented overlooks the efficiency of a Type 1 process by suggesting that Type 2 processes are always engaged (De Neys, 2012). Moreover, the findings we presented showed that validity judgments can complete quickly and are less prone to error; similar to the evidence suggesting that the detection of formal structures automatically trigger basic rule-like schemata (Reverberi et al., 2010). In addition, Morsanyi and Handley (2008) showed that cognitive ability relates to the use of prior knowledge when reasoning which taken together these findings are at odds with the idea that Type 1 processes = associative and Type 2 = rule based.

De Neys (2012) offers a combined PC and DI model as an alternative account (see Figure 5.1)

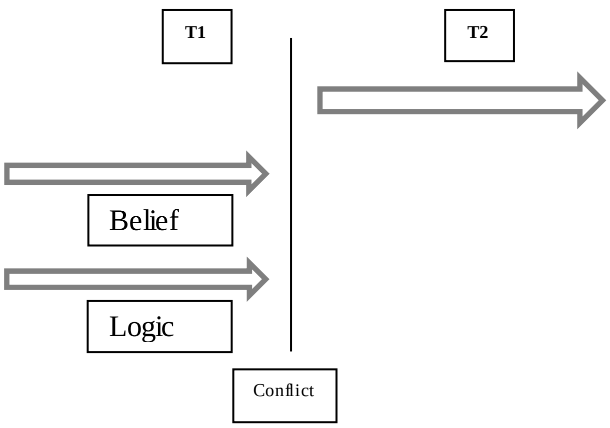


Figure 5.1
A PC-DI model to show Belief and Logic as both intuitive Type 1 processes, followed by deliberate Type 2 processing.

This model shows that deliberative Type 2 processing does not occur until conflict is detected between two competing Type 1 responses. This addresses the inefficiency of Sloman's original PC model and focuses on the idea that logic based judgments can be processed intuitively (see De Neys & Glumicic, 2008; Morsanyi & Handley, 2012). To explain the traditional belief-bias effect, De Neys (2014) argues that whilst both knowledge based and logical processing occur simultaneously at the intuitive level, the `salience' of belief based responses entices participants to give a heuristic response. However, the findings presented in this thesis along with the work by Handley et al.

(2011), have shown that logic can interfere with belief based judgments which would suggest that the logical solution is more `salient'. The model illustrated in Figure 5.1 does not really explain why a belief based response would be more salient than a logical response; additionally it fails to indicate whether deliberative processing could also relate to belief based reasoning.

To reiterate the key findings from the research presented: belief judgments are more prone to error, whilst logic judgments often complete quickly and are less prone to error. When the reasoning task is presented in a blocked design, or when WM demands are increased with an additional instructional condition, logic judgments take longer to complete. However, performance is generally poorer on belief conflict problems and these are more strongly associated with WM demands, inhibition and cognitive style. Taken together these findings suggest that belief based judgments involve effortful Type 2 processing. In the next section an alternative PC dual process interpretation of the findings is presented.

5.4 An Alternative PC Account

In our view and supported by De Neys' work on conflict detection, certain logical inferences can be drawn automatically. We argue that a model that best fits the data under belief instruction is one in which the logical inference is accomplished rapidly and intuitively and requires inhibiting in order to generate a belief based response. However, this model would not explain why RNG interferes with logic judgments or why they take longer to complete in some experiments.

In order to resolve these somewhat contradictory findings, we suggest that certain factors influence logicality. For example, the Feeling of Rightness (Thompson, Prowse

Turner & Pennycook, 2011) generated by factors such as conceptual fluency (Morsanyi and Handley, 2012) can act as a logical cue outside conscious awareness. Whilst factors that can influence explicit logical inferences are linked to WM, intelligence and certain cognitive styles (Macpherson & Stanovich, 2007; Stanovich, 2005; Stanovich & West, 1997).

Therefore the proposal is that there are two routes to a logical solution; the first is the intuitive, fast completing logical route (see De Neys, 2012:2014) based on the feelings associated with valid and invalid arguments. These intuitive feelings about the conclusion trigger the Type 1 output that interferes with the ability give a belief based response. This Type1/Type2 conflict is responsible for the prevalent belief/logic interaction which shows logic having a greater impact on belief judgments. Therefore, we suggest that when instructed to reason on the basis of belief a modified version of a DI model fits the data, where a logical solution is generated intuitively and has to be inhibited until a belief based response is generated (see Figure 5.2).

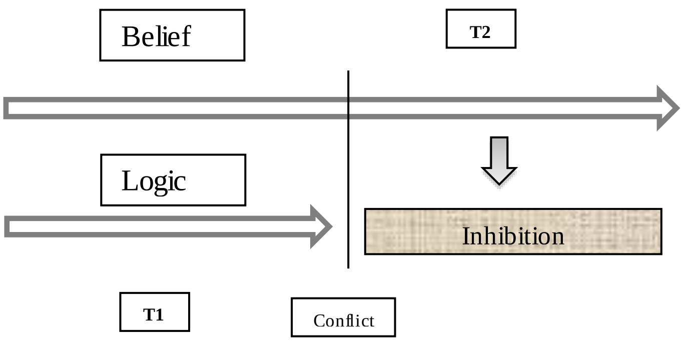


Figure 5.2
A PC-Default Interventionist interpretation of ***Belief based*** *judgments.*

When explicitly instructed to reason logically, the second, deliberative route to a logical solution is called on. This effortful route to a logical solution is impacted on by RNG (Experiment 5), takes longer to complete (Experiments 5, 7, 8 & 9) and is affected by increases in the complexity of the task. However, due to the relatively simple nature of these logical inferences, they are less prone to error. Consequently, a parallel processing model offers an appropriate interpretation of the data when explicitly instructed to reason logically. The model can explain the impact of load on accuracy scores and the increase in latencies (see Figure 5.3).

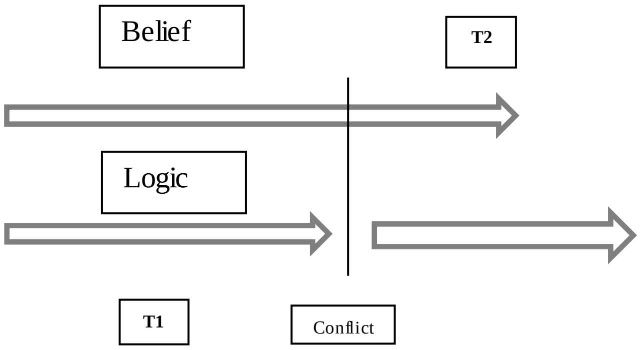


Figure 5.3
*A Parallel Competitive interpretation of **Logic based** reasoning.*

Additional support for the theory that there are two paths to a logical solution is presented in the next section with the examination of behavioural data on problem complexity and individual differences measures.

5.5 Further Support for Two Routes to a Logical Solution

The first route to a logical solution, as presented in Figure 5.2, demonstrates the Type1/Type 2 conflict between intuitive logic and effortful beliefs. We argued that conceptual fluency and a feeling of rightness corresponds with a logical output being more strongly activated and is responsible for the impact of logic on belief based judgments. When provided with explicit instructions to reason logically, the second route is required which can be impacted on by factors such as WM resources. The data on both simple MP and complex disjunctives provides support for the two route theory.

5.5.1 Complexity

There is evidence that certain inferences can be drawn automatically, for example according to Radar and Sloutsky (2002) the direct or automatic route to a logical inference is connected to the conditional `if' or the conjunction `or' which are drawn during discourse processing. Radar and Sloutsky presented participants with short stories without conclusions, which corresponded with the MP or AC (affirming the consequent) inferences. They found that MP inferences primed a conclusion that participants mistakenly thought they had read, suggesting that logical MP inferences can be drawn without awareness (also see Leo & Greene, 2008; Reverberi, Pischedda, Burigo & Cherubini, 2012). The experimental data examining problem complexity in this thesis has shown that that conflict between belief and logic is greater for MP arguments (Experiment 4, 7 8 and 9). This supports the idea that the intuitive route to a logical solution completes first and requires inhibition if instructed to reason on the basis of beliefs.

In contrast, these findings also showed that the effect of conflict was much smaller for disjunctive judgments which can be regarded as support for the second and explicit route to a logical solution. Figure 5.4 illustrates how the conflict between belief and logic occurs much later, during Type 2 processing. When instructed to reason on the basis of belief, inhibition of the logic response may still be required but to a lesser extent, which results in less conflict between the two responses.

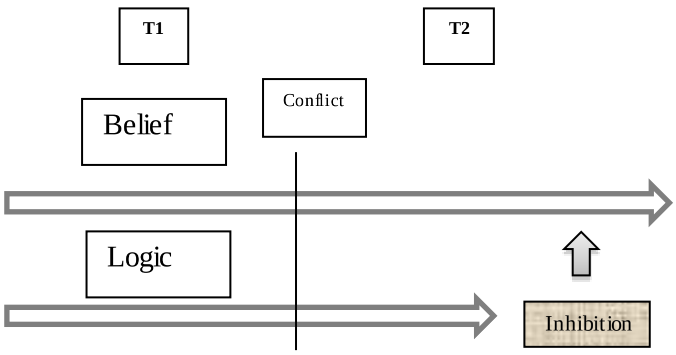


Figure 5.4
A PC model illustrating the Belief/Logic conflict for Disjunctive arguments. These problems require Type 2 processing and conflict occurs much later.

In Experiment 5, 7, 8 & 9, logic judgments look longer to complete and in Experiment 9 there was an impact of conflict on logic latencies in the experimental condition. As further support for the second route to a logical solution, we conjectured that the parallel processing of belief judgments whilst answering according to logic, impacted on the time it took to produce an explicit logical output. This effect along with the small difference between logic conflict and no-conflict problems suggests that belief based responses can occasionally interfere with a logical output. We hypothesise that the level of conflict will depend on the complexity of the argument. For example, consider this more complex syllogistic problem:

All doctors are fishermen

Some fishermen are violinists

Therefore, some doctors are not violinists

For these types of problems a belief based response would complete first and require inhibiting in order for a logical output to complete. According to the model illustrated in Figure 5.5 both belief and logic based responses demand Type 2 processes creating a Type 2/Type 2 conflict. Unlike simple inferences there is less likely to be an intuitive route to a logical solution but, depending on the complexity of the syllogism, a belief based response may be available more rapidly.

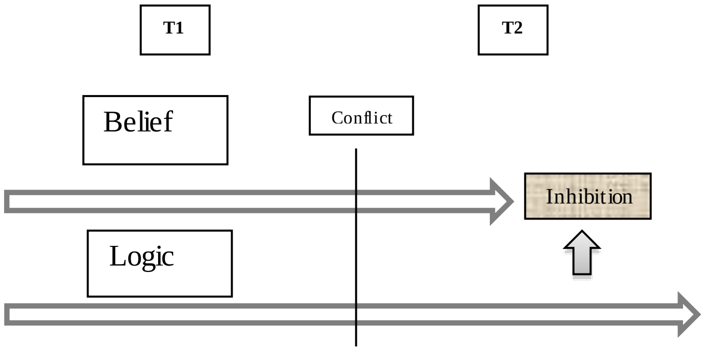


***Figure* 5.5**
PC model illustrating the Belief/Logic conflict for complex Syllogistic arguments.

5.5.2 Individual Differences

Consistent with the data that shows the CRT relating to logic based reasoning in Experiment 4 and belief based reasoning in Experiment 7 and 9, this could also provide support for two routes to a logical solution. Low CRT scorers are characterised as those more inclined to give a reflexive, intuitive response to a question that requires some deliberation to arrive at the correct answer. In Experiment 7 and 9 the impact of conflict was greater on beliefs for low CRT scorers than high CRT scorers, suggesting low CRT scorers rely more on their intuitive logical cues which results in more interference with belief based judgments (refer to Figure 5.2).

When instructed to engage in explicit logical reasoning this requires effortful Type 2 processing which will be influenced by factors such as WM and intelligence. Although the effect is a weak one, the relationship between CRT, AH4 and logic based inferences in Experiment 4 is consistent with a second and explicit route to a logical solution.

5.4.2 Summary of the Findings

The research presented in this thesis suggests that accessing prior knowledge to formulate a conclusion requires effortful processing that is prone to errors. When instructed to reason about the validity of an argument, the integration of premise information and model construction also requires effortful processing which can be impacted on by a secondary load and take longer to complete. In contrast, simple arguments can cue an intuitive logical response, accompanied by a feeling of rightness associated with the validity of the argument, which cues a Type 1 output that interferes with belief based judgments. With more complex arguments such as syllogisms, a belief based response may complete first impacting on a logic based response. The data presented fits with a PC account that suggests both the logical structure of an argument and relevant knowledge are processed simultaneously. However, the direction of conflict experienced by the reasoner will depend on factors such as the complexity of the problem, the instructions delivered and the ability to draw on the appropriate executive resources for Type 2 processing. Both background knowledge and the structure of a problem provide important information that can be adhered to or dismissed and conflict between these sources of information can arise at the intuitive (Pennycook, Trippas, Handley & Thompson, 2013) or the deliberative (Banks & Hope, 2014) level of processing.

5.6 Research on Biases

It is important to emphasise that the standard belief-bias effect in reasoning occurs in the context of instructing people to reason logically. Failure to do so inevitably leads to a belief based judgment since it is the only alternative response available in face of WM load (De Neys, 2006), cognitive ability and WM capacity (Macpherson & Stanovich, 2007; Stanovich, 2005; Stanovich & West, 1997; Stanovich &West, 2000). Correlations with performance on reasoning tasks may have nothing to do with belief-bias but might have everything to do with these ID measures reflecting an individual's capacity to engage in the task in line with the instruction presented. The issue that arises from this is that our understanding of a biased reasoner or decision maker is only relative to the normative standard of the instructions provided which differ from one type of problem to the next.

Research concerned with manipulating instruction has revealed different patterns of belief-bias which is not only the case in terms of syllogisms (Evans, Newstead, Allen & Pollard, 1994) but has been shown to affect the Conjunction Fallacy and Base Rate problems. For example, the Conjunction Fallacy (CF) or the Linda problem (Tversky & Kahneman, 1983) is interpreted as the use of a Type 1- representative heuristics that leads to the neglect of the conjunction rule which is a basic rule of probability. This representative bias disappears when the problem is rephrased in terms frequencies. The same effect occurs for base rate problems (Gigerenzer, 1991), which causes us to question whether the focus of the problem or the way in which people are instructed is responsible for the outcome. For example does focusing on a specific person (Linda) in the Conjunction problem encourage a representative response, whilst the frequency format encourages a more general `normative' response? Additionally, are the responses to the instruction `which is more probable?' against `how many are there?' really indicative of the same `normative standard' or even measuring the same processes required to make these fundamentally different types of judgment.

The processing of base-rate information is often regarded as requiring Type 2 processing, however, Pennycook et al. (2013) found when they asked participants to respond to base-rate problems according to beliefs or statistical information, both sources of information where available at the Type 1 level. They found that participants where just as fast and as confident about making judgments based on statistics as they were about judgments based on beliefs and these interfered with each other to the same extent. Therefore, ignoring base-rate information does not necessarily mean the information is not processed on some level.

Similarly with the CRT, a normatively incorrect response on this measure does not necessarily mean that the 'intuitive' response is based on Type 1 processes. Perhaps the CRT is not about intuitive responding but about less processing being required for one type of response. For example, an incorrect response on any of the CRT questions is attributed to an erroneous intuitive (Type 1) output. However, even the incorrect response would involve some simple arithmetic to arrive at that answer. For example, take the third problem from the CRT;

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

The majority of people who answer incorrectly say 24; to arrive at 24 you would need to carry out the simple maths of dividing 48 by 2. If this simple processing was not occurring and people were not doing some mental arithmetic then there would not be a standard incorrect answer. The number 24 would be given as frequently as any other number and this is not the case. It is quite possible that the response regarded as intuitive is actually an effortful response that simply completes first producing a Type 2/Type 2 conflict. The CRT is a really good example of a problem where, like belief-bias, there is a simple cue to an erroneous solution. However, the fact that the cue is readily available does not necessarily mean that the response is driven by Type 1 processing. Perhaps what should be focused on is the role of inhibition as a way of clarifying the type of conflict that is occurring in many of these problems?

The assumptions and models offered in this thesis are relative to the instructions given to the participants and are specific to the type of problems used. Although some predictions are made about the outcomes using more complex syllogisms an important implication for research on biases is to not be tempted to make inferences about the state of response based on the fact that its activated rapidly, linked to limited cognitive ability or incorrect from a normative perspective because this does not necessarily mean it is intuitive in nature. Any changes to the direction of conflict and levels of processing will depend on problem type, instructions and the executive processes involved (i.e. inhibition).

5.7 Future Research

One of the key issues in determining the direction of interference between logic and belief judgments based on the models presented in section 5.5 is the complexity of the argument. As discussed earlier one would expect the impact of conflict to be greater on logical judgments with more complex arguments and smaller with belief judgments. An interesting and potential area of development is to manipulate the complexity of beliefs by using materials that vary in terms of degree of belief and examine the impact on logic based reasoning as well as measuring how logical processing compromises beliefs. One might expect that with beliefs that are more complex to evaluate the impact of conflicting logical structure will be greater.

In the last experimental chapter of this thesis we conjectured that inhibition plays a key role in dealing with conflict between belief and logic based reasoning, therefore another focus for future research could involve the manipulation of inhibition. There have been several studies to show a reduction in the size of colour–word interference through training on the Stroop task (Davidson, Zacks & Williams, 2003; Macleod, 1998), along with improved response inhibition and reasoning through working memory training in children with ADHD (Klingberg et al., 2005). Further research on the effects of inhibition training has been carried out on the "Matching Bias" (Moutier, Angeard & Houdé, 2002; Houdé & Moutier, 1999) and the "Conjunction Fallacy" tasks (Cassotti & Moutier, 2010). The training strategies used logical explanation and inhibition training which included emotional warnings about the risk of error. In both cases, logical explanation was not sufficient at overriding the heuristic response but inhibition training with emotional warning drastically changed the cortical anatomy of reasoning as a specific result of executive training (Houdé et al., 2000). It would be interesting to determine whether inhibition could be "trained" in order to modify some of findings in the research presented in this thesis. On the understanding that executive training aimed directly at bias inhibition can change reasoning performance and lead to neural reconfigurations (Moutier, Plagne-Cayeux, Melot & Houdé, 2006), perhaps inhibition training would improve an individual's ability to inhibit the fast completing logic responses for simple arguments. Conversely, similar training using more complex inferences may improve the inhibition of faster completing belief responses.

Developmental research on inhibition has shown that its capacity produces a curvilinear age trend. The ability to resist prepotent responses tends to improve from childhood to adolescence and declines again in later life (Bedard et al., 2002; Christ, White, Mandernach & Keys, 2001). De Neys & Van Geldor (2008) demonstrated that consistent with inhibitory capacities, syllogistic reasoning performance produced the same curvilinear pattern. Specifically they showed that when belief and logic conflict, performance is determined by a person's aptitude for inhibiting belief based responses.

Based on the research presented in this thesis one would expect that the effect of conflict on belief judgments to show a similar curvilinear trend. However the WM demand of the task would also depend on the type of arguments used. For example, Modus Ponens inferences can be drawn relatively automatically (Reverberi et al. 2012) whilst more complex inferences such as Modus Tollens are available later (Reverberi, Burgio & Cherubini, 2009). One would expect no effect of conflict if inhibition resources were underdeveloped (in young children) or limited in any way.

5.8 Conclusion

This thesis has examined the impact of a secondary load on simple judgments using the novel methodology introduced by Handley et al. which instructed participants to reasoning on both the basis of logic and beliefs. The role of WM and inhibition in relation to both reasoning types was also investigated, using original variations of a Stroop and *n*-back task along with several ID measures. The main objective of this research was to advance the debate concerning DPT in reasoning, moving away from an exclusively DI model and toward a version of the PC approach to dual processing in an attempt to interpret our findings. We argued that the data was consistent with there being two routes to a logical solution for simple judgments; one intuitive and one deliberative. The findings also confirmed that effortful belief based judgments demand on executive resources; however, both background knowledge and the structure of a problem can provide important information that can be followed or disregarded. Finally, the conflict experienced between knowledge based and structural information can occur at a Type1/Type 2 level or a Type 2/Type 2 level.

In this thesis, we have shown a function for intuitive logic and support for the effortful nature of belief judgments. Future research should focus on the important role instructions, problem type, complexity and executive resources play in our interpretation of research on biases in human reasoning.

References

Allport, A., & Wylie, G. (2000). Task switching, stimulus-response bindings, and negative priming. *Control of cognitive processes: Attention and performance XVIII*, 35-70.

Bacon, A., Handley, S., & Newstead, S. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning, 9*(2), 133-168.

Baddeley, A. D. (1966). The capacity for generating information by randomization. *The Quarterly Journal of Experimental Psychology, 18*(2), 119-129.

Baddeley, A. D. (1986). Working Memory. *Oxford; Oxford University Press*.

Baddeley, A. D. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology: Section A, 51*(4), 819-852.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology, 36*(2), 233-252.

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning:eye-movement evidence for selective processing models. *Experimental Psychology 53*(1), 77-86.

Banks, A. P., & Hope, C. (2014). Heuristic and analytic processes in reasoning: An event-related potential study of belief bias. *Psychophysiology, 51*(3), 290-297.

Barrouillet, P., & Lecas, J.-F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*(4), 289-302.

Bedard, A.-C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental neuropsychology, 21*(1), 93-111.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 12-23.

Biek, M., Wood, W., & Chaiken, S. (1996). Working knowledge, cognitive processing, and attitudes: On the determinants of bias. *Personality and Social Psychology Bulletin, 22*(6), 547-556.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives. Vol. 1: Cognitive Domain*. New York: David McKay Company.

Brown, R., Soliveri, P., & Jahanshahi, M. (1998). Executive processes in Parkinsons disease—random number generation and response suppression. *Neuropsychologia, 36*(12), 1355-1362.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning, 9*(3), 203-244.

Cassotti, M., & Moutier, S. (2010). How to explain receptivity to conjunction-fallacy inhibition training: Evidence from the Iowa Gambling Task. *Brain and Cognition, 72*(3), 378-384.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: The Guilford Press.

Christ, S. E., White, D. A., Mandernach, T., & Keys, B. A. (2001). Inhibitory control across the life span. *Developmental neuropsychology, 20*(3), 653-669.

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making* (pp. 1-40). Hove: Psychology Press.

Colom, R., Abad, F. J., Quiroga, M., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence, 36*(6), 584-606.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences, 7*(12), 547-552.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual. *Odessa, FL: Psychological Assessment Resources*.

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Aging, Neuropsychology, and Cognition, 10*(2), 85-98.

De Neys, W. (2006). Dual Processing in Reasoning -Two Systems but One Reasoner. *Psychological Science, 17*(5), 428-433.

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science, 7*(1), 28-38.

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*(2), 169-187.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PloS ONE, 6*(1), e15954. doi: 15910.11371/journal.pone.0015954.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*(1), 45-61.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*(3), 1248-1299.

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 208-216.

De Neys, W., Novitskiy, N., Ramautar, J., & Wagemans, J. (2010). *What makes a good reasoner?: Brain potentials and heuristic bias susceptibility*. Paper presented at the The Annual Conference of the Cognitive Science Society (Vol. 32, pp.1020-1025).

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology 54*(2), 128-133.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning, 11*(2), 123-150.

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: the rise and fall of belief inhibition during syllogistic reasoning. *Developmental science, 12*(1), 123-130.

De Neys, W. D., Vartanian, O., & Goel, V. (2008). Smarter Than We Think: When our brains detect that we are biased. *Psychological Science, 19*(5), 483-489.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptalizing change in the cognitive construction of knowledge. *Educational psychologist, 33*(2-3), 109-128.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), Models of working memory: Mechanisms of active maintenance and executive control (pp. 102-134). Cambridge: Cambridge University Press.

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

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709-724.

Epstein, S., & Meier, P. (1989). Constructive thinking: A broad coping variable with specific components. *Journal of Personality and Social Psychology, 57*(2), 332-350.

Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *The Quarterly Journal of Experimental Psychology, 24*(2), 193-199.

Evans, J. St. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning, 4*(1), 45-110.

Evans, J. St. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences, 7*(10), 454-459.

Evans, J. St. B. T. (2006). *Dual system theories of cognition: Some issues.* Paper presented at the The 28th Annual Meeting of the Cognitive Science Society (pp. 202-207).

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

Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339.

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

Evans, J. St. B. T. (2009). How many dual-process theories do we need? One, two, or many? In K. Frankish (Ed.), *In Two Minds: Dual Processes and Beyond* (pp. 33-54). New York: Oxford University Press.

Evans, J. St. B. T. (2010a). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.

Evans, J. St. B. T. (2010b). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*(4), 313-326.

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*(2–3), 86-102. doi: 10.1016/j.dr.2011.07.007

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition, 11*(3), 295-306.

Evans, J. St. B. T., & Brooks, P. G. (1981). Competing with reasoning: A test of the working memory hypothesis. *Current Psychology, 1*(2), 139-147.

Evans, J. St. 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. St. B. T., Handley, S. J., & Bacon, A. M. (2009). Reasoning Under Time Pressure. *Experimental Psychology, 56*(2), 77-83. doi: 10.1027/1618-3169.56.2.77

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology, 63*(5), 892-909.

Evans, J. St. B. T., Newstead, S., Allen, J., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology, 6*(3), 263-285.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and Reasoning*. Hove: Psychology Press.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223-241.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54*(1), 1-71.

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

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101-135.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science, 17*(2), 172-179. doi: 10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*(2), 201-225.

Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2014). Feelings of error in reasoning–in search of a phenomenon. *Thinking & Reasoning*, 1-14. doi: 10.1080/13546783.2014.980755

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky. *Psychoogical Review 103*(3), 592-596.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review, 103*(4), 650-669.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin, 119*(1), 23-26.

Gilhooly, K., Logie, R., Wetherick, N., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & cognition, 21*(1), 115-124.

Gilhooly, K., Logie, R., & Wynn, V. (2002). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology, 21*(2), 111-120.

Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and aging, 9*(3), 356-371.

Glahn, D. C., Ragland, J. D., Abramoff, A., Barrett, J., Laird, A. R., Bearden, C. E., & Velligan, D. I. (2005). Beyond hypofrontality: A quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Human brain mapping*, *25*(1), 60-69.

Goel, V. (2003). Evidence for dual neural pathways for syllogistic reasoning. *Psychologica, 32*, 301-309.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage, 12*(5), 504-514.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*(1), 11-22.

Goel, V., & Vartanian, O. (2011). Negative emotions can attenuate the influence of beliefs on logical reasoning. *Cognition and Emotion, 25*(1), 121-131.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*(2), 175-195. doi: 10.1080/13546780442000051

Handley, S. J., Capon, A., Copp, C., & Harper, C. (2002). Conditional reasoning and the Tower of Hanoi: The role of spatial and verbal working memory. *British Journal of Psychology, 93*(4), 501-518.

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 28-34.

Heim, A. (Ed.). (1970). *AH4 group test of general intelligence manual*. London: Nelson.

Horne, R. L., Evans, F. J., & Orne, M. T. (1982). Random number generation, psychopathology, and therapeutic change. *Archives of General Psychiatry, 39*(6), 680-683.

Houdé, O., & Moutier, S. (1999). Deductive reasoning and experimental inhibition training: The case of the matching bias. New data and reply to Girotto. *Cahiers de psychologie cognitive, 18*(1), 75-85.

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of cognitive neuroscience, 12*(5), 721-728.

Jahanshahi, M., Profice, P., Brown, R. G., Ridding, M. C., Dirnberger, G., & Rothwell, J. C. (1998). The effects of transcranial magnetic stimulation over the dorsolateral prefrontal cortex on suppression of habitual counting during random number generation. *Brain, 121*(8), 1533-1544.

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological review, 99*(3), 418-439.

Jonides, J., & Smith, E. E. (1997). The architecture of working memory. In M. D. Rugg (Ed.), Cognitive Neuroscience (pp. 243-276). Cambridge, MA: MIT Press.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 49-81). Cambridge: Cambridge Universirty Press.

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science, 4*(6), 533-550.

Kim, S. Y., Kim, M. S., & Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proceedings of the national Academy of Sciences of the United States of America, 102*(45), 16524-16529.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology, 55*(4), 352-358.

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.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological review, 107*(4), 852-884.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., . . . Westerberg, H. (2005). Computerized training of working memory in children with ADHD-a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(2), 177-186.

Kokis, J. V., 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 styles. *Journal of experimental child psychology, 83*(1), 26-52.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological review, 118*(1), 97-109.

Lea, R. B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1469-1482.

Leo, P. D., & Greene, A. J. (2008). Is awareness necessary for true inference? *Memory & cognition, 36*(6), 1079-1086.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology, 38*(2), 229-247.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove: Lawrence Erlbaum Associates LTD.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In R. H. Logie & M. Denis (Eds.), Mental Images in Human Cognition (Vol. 80, pp. 105-115). Amsterdam: Elsevier Science Publishing Company, INC.

Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences, 17*(2), 115-127.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22*(1), 5-36.

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking & Reasoning, 8*(2), 97-107.

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta psychologica, 106*(3), 303-327.

Milligan, G. W. (1987). The use of the arc-sine transformation in the analysis of variance. *Educational and Psychological Measurement, 47*(3), 563-573.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences, 6*, 414-417.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive psychology, 41*(1), 49-100.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621-640.

Monsell, S. (2003). Task switching. *Trends in cognitive sciences, 7*(3), 134-140.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*(2), 111-121.

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of experimental child psychology, 99*(1), 18-36.

Morsanyi, K., & Handley, S. J. (2012). Logic feels so good—I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 596-616.

Moutier, S., Angeard, N., & Houdé, O. (2002). Deductive reasoning and matching-bias inhibition training: Evidence from a debiasing paradigm. *Thinking & Reasoning, 8*(3), 205-224.

Moutier, S., & Houdé, O. (2003). Judgment under uncertainty and conjunction fallacy inhibition training. *Thinking & Reasoning, 9*(3), 185-201.

Moutier, S., Plagne-Cayeux, S., Melot, A.-M., & Houdé, O. (2006). Syllogistic reasoning and belief-bias inhibition in school children: evidence from a negative priming paradigm. *Developmental science, 9*(2), 166-172.

Newstead, S. E. (2000). Are there two different types of thinking? *Behavioral and Brain Sciences, 23*(05), 690-691.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology Section A, 57*(1), 33-60.

Osman, M. (2002). *Is there evidence for unconscious reasoning processes.* Paper presented at the 24th Annual Meeting of the Cognitive Science Society, Washington, DC.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*(6), 988-1010.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping, 25*(1), 46-59.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*(6), 972-987.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*(2), 307-317.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition, 123*(3), 335-346.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2013). Base rates: both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 544-554.

Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory, 7*(2), 209-231.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology: Section A, 53*(4), 1202-1223.

Rabbitt, P. (1997). *Methodology of frontal and executive function*. Hove: Psychology Press.

Rader, A. W., & Sloutsky, V. M. (2002). Processing of logically valid and logically invalid conditional inferences in discourse comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 59-68.

Raz, A., Shapiro, T., Fan, J., & Posner, M. I. (2002). Hypnotic suggestion and the modulation of Stroop interference. *Archives of General Psychiatry, 59*(12), 1155-1161.

Reber, A. S. (1996). *Implicit learning and tacit knowledge: An essay on the cognitive unconcious*. New York, NY: Oxford University Press.

Reverberi, C., Burigo, M., & Cherubini, P. (2009). A subliminal premise can automatically trigger an elementary deductive inference. *Bicocca Open Archive*, 201-205.

Reverberi, C., Cherubini, P., Frackowiak, R. S., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human brain mapping, 31*(9), 1430-1445.

Reverberi, C., Pischedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta psychologica, 139*(1), 244-253.

Rips, L. J. (1988). Deduction In R. J. Sternberg & E. E. Smith (Eds.), The psychology of human thought (pp. 116-152). New York, NY: Cambridge University Press.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: The MIT Press.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*(3), 497-510.

Schmiedek, F., Li, S.-C., & Lindenberger, U. (2009). Interference and facilitation in spatial working memory: age-associated differences in lure effects in the n-back paradigm. *Psychology and aging, 24*(1), 203-210.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review, 84*(2), 127-190.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental psychology, 39*(6), 964-975.

Singmann, H., Klauer, K. C., & Kellen, D. (2014). Intuitive Logic Revisited: New Data and a Bayesian Mixed Model Meta-Analysis. *PloS one, 9*(4), e94223. doi: 10.1371/journal.pone.0094223

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin, 119*(1), 3-22.

Stanovich, K. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Stanovich, K. E. (2005). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In J. St. B. Evans & K. Frankis (Eds.), In Two Minds: Dual Processes and Beyond (pp. 55-88). Oxford: Oxford University Press.

Stanovich, K. E. (2009b). *What intelligence tests miss: The psychology of rational thought*. New Haven, CT: Yale University Press.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society, 11*(1), 3-13.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342-357.

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

Stanovich, K. E., & West, R. F. (1998b). Cognitive ability and variation in selection task performance. *Thinking & Reasoning, 4*(3), 193-230.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645-665.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning, 13*(3), 225-247.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672-695.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Stupple, E. J., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning, 14*(2), 168-181.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive psychology, 63*(3), 107-140.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology, 46*(4), 679-699.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition, 39*(7), 1275-1289.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147-168.

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, 30*(4), 583-591.

Towse, J. N., & Valentine, J. D. (1997). Random generation of numbers: A search for underlying processes. *European Journal of Cognitive Psychology, 9*(4), 381-400.

Tsujii, T., Masuda, S., Akiyama, T., & Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia, 48*(7), 2005-2008.

Tsujii, T., & Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: A near-infrared spectroscopy study. *Brain research, 1287*, 118-125.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

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

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

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11*(3), 239-278.

Verschueren, N., Schaeken, W., & d’Ydewalle, G. (2004). *Everyday conditional reasoning with working memory preload.* Paper presented at the The 26th annual meeting of the Cognitive Science Society (pp. 1399-1404).

Wason, P., & Evans, J. (1975). Dual processes in reasoning? *Cognition, 3*(2), 141-154.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), New Horizons in Psychology (Vol. 1, pp. 135-151). Harmondsworth: Penguin Books.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology, 20*(3), 273-281.

Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of psychology, 102*.

Appendices

Appendix A: Brief, Instructions & Debrief

Experiment 1: Brief, Instructions, Debrief

Brief & Consent Form

In this Experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If a person is swimming, then they are dry

Suppose and person is swimming

Is the person wet?

On some trials you will be asked to judge whether the conclusion follows logically and on other trials you will be asked whether the conclusion is believable.

Whilst making these judgments you will also be asked to remember some information about the location of a series of dots in a grid. (See instructions for details). The whole procedure will take approximately 25-35 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

I understand the aims of this research; that I am free to withdraw from the research at any time and ask for my data to be destroyed if I wish (for up to two weeks after the experiment was run) and that my data will remain anonymous.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities.

Under these circumstances, I agree to participate in the research.

Name: ……………………………………….

Signature: ....…………….. Date: ....………….

Instructions

The instructions were delivered via E-prime as follows:

In the following experiment you will be presented with 64 reasoning problems of the following type:

If a child is crying, then it is sad

Suppose a child is crying

Does it follow that the child is sad?

For each problem you will have to answer either according to your beliefs or according to logic. When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If you finish a drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be full?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that if you finish your drink then the glass will be empty, therefore the conclusion is unbelievable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentences. A valid conclusion is one that logically follows.

For example:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass is empty?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that "if you finish your drink then the glass will be full" and supposing you "finish your drink" you must logically conclude that your glass will be full. This is why the conclusion "does it follow that the glass is empty" is logically invalid.

For every problem you will first be presented with one statement. Once you have read this statement press space to see the second statement, the conclusion and response options. At this point the initial statement will disappear.

The response options will either show valid/invalid or believable/unbelievable. If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs.

Please indicate your choice by pressing: s) for VALID or BELIEVABLE and k) for INVALID or UNBELIEVABLE; depending on the available options.

Before every belief/logic problem you will be presented with a pattern of dots in a 3 x 3 grid for a very short time. It is your task to remember this pattern before solving the main problem (which you will be presented with after studying the pattern).

After responding to the main problem you will be presented with an empty grid in which you have to reproduce the dot pattern by clicking in the empty spaces.

If you make a mistake, click the dot again to make it disappear. After reproducing the entire dot pattern, click `confirm' to move on to the next problem.

Make sure you reproduce the dot pattern correctly, or the trial becomes invalid. However, try to respond as quickly and accurately as possible on both tasks.

You will now be presented with 8 trials to practice these instructions. These do not count towards your total performance.

For the first four trials, you will be presented with four belief/logic reasoning problems. In the final four trials, you will be presented with the dot pattern task alongside the belief/logic reasoning problems.

If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

You will now be presented with the last 4 practice trials. Please remember the dot pattern, make the belief or logic judgment and correctly reproduce the pattern using the mouse.

This concludes the practice trials.

If you have any additional questions, please ask them now.

If not PRESS SPACE TO BEGIN WITH THE ACTUAL EXPERIMENT

Debrief

Anonymous participant ID: ________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument.

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

The aim of this research is to evaluate the degree to which logic and beliefs interfere with one another when participants are asked to make judgments based upon either the structure of an argument or knowledge relating to its content.

To test the hypothesis that reasoning on the basis of beliefs requires effortful processing; we added a secondary task to the Experiment. Participants were initially assigned to High and Low load conditions of the secondary task, which consisted of a Dot Memory Task being presented to the participant prior to the appearance of the problem and then asked to recall the pattern after they had made their inference. This task has been shown to interfere with conditional reasoning under logic instructions. We believe that if the secondary task interferes with belief based instructions in this research, then belief based reasoning requires more effort than previously reported.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below:

Researcher:	Stephanie Howarth:
	stephanie.howarth@plymouth.ac.uk
Supervisor:	Professor Simon Handley:
	shandley@plymouth.ac.uk
Second Supervisor:	Dr Clare Walsh

clare.walsh@plymouth.ac.uk

Alternative Contact: Paula Simson: Secretary to the Human Ethics Committee

paula.simson@plymouth.ac.uk

Contact Tel: 01752 584503

References:

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. Trends in Cognitive Sciences, 7, 454-459

Handley, S.J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs and instruction: A test of the default interventionist account of belief bias. Journal of Experimental Psychology: Learning, Memory and Cognition, 37,28-43.

Experiment 2: Brief, Instructions, Debrief

Brief

For the Random Number Generation Group in Experiments 2 & 3:

In this Experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If a person is swimming, then they are dry

Suppose a person is swimming

Is the person wet?

On some trials you will be asked to judge whether the conclusion follows logically and on other trials you will be asked whether the conclusion is believable.

Whilst making these judgments you will also be asked to say out aloud a random number between 1 and 9, every second, on the beep of a metronome. (See instructions for details). The whole procedure will take approximately 20-30 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

For the Articulatory Suppression Group in Experiment 2 only:

In this Experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If a person is swimming, then they are dry

Suppose a person is swimming

Is the person wet?

On some trials you will be asked to judge whether the conclusion follows logically and on other trials you will be asked whether the conclusion is believable.

Whilst making these judgments you will also be asked to say aloud the number two, every second, on the beep of a metronome. (See instructions for details). The whole procedure will take approximately 20-30 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

For the Control Group in Experiments 2 & 3:

In this Experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If a person is swimming, then they are dry

Suppose a person is swimming

Is the person wet?

On some trials you will be asked to judge whether the conclusion follows logically and on other trials you will be asked whether the conclusion is believable.

The whole procedure will take approximately 15-20 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

Consent form

The same form was used across each group in experiments 2 – 5 & 7 – 9.

I understand the aims of this research; that I am free to withdraw from the research at any time and ask for my data to be destroyed if I wish (for up to two weeks after the experiment was run) and that my data will remain anonymous.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities.

Under these circumstances, I agree to participate in the research.

Name: ……………………………………….

Signature: ....…………….. Date: ....…………..

Instructions

For the <u>Random Number Generation</u> Group:

In the following experiment you will be presented with 64 reasoning problems of the following type:

If a child is crying, then it is sad

Suppose a child is crying

Does it follow that the child is sad?

For each problem you will have to answer either according to your beliefs or according to logic. When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If you finish a drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be full?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that if you finish your drink then the glass will be empty, therefore the conclusion is unbelievable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentences. A valid conclusion is one that logically follows.

For example:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass is empty?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid. In this particular example the correct answer according to logic is INVALID, because the first premise states that "if you finish your drink then the glass will be full" and supposing you "finish your drink" you must logically conclude that your glass will be full. This is why the conclusion "does it follow that the glass is empty" is logically invalid.

For every problem you will first be presented with one statement. Once you have read this statement press space to see the second statement, the conclusion and response options. At this point the initial statement will disappear. The response options will either show valid/invalid or believable/unbelievable.

If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs.

Please indicate your choice by pressing s) for VALID or BELIEVABLE and k) for INVALID or UNBELIEVABLE; depending on the available options.

During this experiment a metronome beep will sound every second (1000ms). This will be audible through your earphones. It is your task to say aloud a number between 1-9 on each beep; for the entire duration of the experiment; and your responses will be recorded on a dicta-phone. You are to continue generating random numbers whilst solving the reasoning problems.

In order to understand the concept of randomness, please consider the following:

Suppose you had written the numbers 1-9 on pieces of paper and put them in a hat. You take out one piece of paper, call out the number and return it to the hat. Then you reach *for another piece of paper and do the same thing. The series of numbers you call out in that way should be random.*

It is important for you to maintain a consistent response rhythm to the generation of numbers. If you should fall behind, please listen to the beep and promptly re-establish your pace.

Place the earphones on now and you will hear the metronome beep. Please practice generating random numbers on each beep for the next 10 seconds.

Keeping your headphones on, you will now be presented with four practice reasoning problems, together with the metronome beep to practice the two together. These practice trials do not count towards your total performance.

If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

You will now be presented with the four practice reasoning trials and metronome beep together.

This concludes the practice trials.

If you have any additional questions, please ask them now.

If not Please PAUSE counting for 10 beeps:

NOW COMMENCE COUNTING AND PRESS SPACE TO BEGIN WITH THE ACTUAL EXPERIMENT

Specific instructions for the Articulatory Suppression Group:

During this experiment a metronome beep will sound every 1000ms. This will be audible through your earphones. It is your task to say aloud the number 2 on each beep, for the entire duration of the experiment, and your responses will be recorded on a dicta-phone.

You are to continue saying 2 on each beep whilst solving the reasoning problems. It is important for you to maintain a consistent response rhythm. If you should fall behind, please listen to the beep and promptly re-establish your pace.

For the Control Group, the instructions are the same as the RNG and AS group minus secondary task instructions.

Debrief

Presented to each group:

Anonymous participant ID: ________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument.

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

The aim of this research is to evaluate the degree to which logic and beliefs interfere with one another when participants are asked to make judgments based upon either the structure of an argument or knowledge relating to its content.

To test the hypothesis that reasoning on the basis of beliefs requires effortful processing; we added a secondary task to the Experiment. Participants were initially assigned to High load, Low load, or Control condition. The High Load condition required participants to randomly generate a number between 1 and 9, every second, whilst completing the reasoning questions. This secondary task is meant to heavily impact on working memory and we believe if this interferes with belief based instructions in this research, then belief based reasoning requires more effort than previously reported. The Low load condition required participants to say the number two out aloud, every second, whilst completing the reasoning questions. Previous research has indicated that this form of Articulatory suppression should not impact greatly on reasoning ability. Finally the control group had no secondary load and completed the reasoning questions alone.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below (see experiment 1)

Experiment 3: Instructions, Debrief

Instructions

For the RNG group:

In the following experiment you will be presented with 64 reasoning problems of the following type:

If a child is crying, then it is sad

Suppose a child is crying

Does it follow that the child is sad?

For each problem you will have to answer either according to your beliefs or according to logic. When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If you finish a drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be full?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that if you finish your drink then the glass will be empty, therefore the conclusion is unbelievable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentences. A valid conclusion is one that logically follows.

For example:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass is empty?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid. In this particular example the correct answer according to logic is INVALID, because the first premise states that “if you finish your drink then the glass will be full” and supposing you “finish your drink” you must logically conclude that your glass will be full. This is why the conclusion “does it follow that the glass is empty” is logically invalid.

For every problem you will first be presented with one statement. You have a few seconds to read the first statement before the second statement, the conclusion and response options appear on the screen.

The response options will either show valid/invalid or believable/unbelievable. If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs.

Please indicate your choice by pressing s) for VALID or BELIEVABLE and k) for INVALID or UNBELIEVABLE; depending on the available options.

During this experiment a metronome beep will sound every second (1000ms). This will be audible through your earphones. It is your task to say aloud a number between 1-9 on each beep; for the entire duration of the experiment; and your responses will be recorded on a dicta-phone.

You are to continue generating random numbers whilst solving the reasoning problems.

In order to understand the concept of randomness, please consider the following:

Suppose you had written the numbers 1-9 on pieces of paper and put them in a hat. You take out one piece of paper, call out the number and return it to the hat. Then you reach for another piece of paper and do the same thing. The series of numbers you call out in that way should be random.

It is important for you to maintain a consistent response rhythm to the generation of numbers. If you should fall behind, please listen to the beep and promptly re-establish your pace.

Place the earphones on now and you will hear the metronome beep. Please practice generating random numbers on each beep for the next 10 seconds.

Keeping your headphones on, you will now be presented with four practice reasoning problems, together with the metronome beep to practice the two together. These practice trials do not count towards your total performance.

If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

You will now be presented with the four practice reasoning trials and metronome beep together.

This concludes the practice trials. If you have any additional questions, please ask them now.

If not please PAUSE counting for 10 beeps.

NOW COMMENCE COUNTING AND PRESS SPACE TO BEGIN WITH THE ACTUAL EXPERIMENT.

For the Control Group, the instructions are the same as above minus secondary task instructions.

Debrief

Presented to both groups:

Anonymous participant ID: ___________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument.

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

The aim of this research is to evaluate the degree to which logic and beliefs interfere with one another when participants are asked to make judgments based upon either the structure of an argument or knowledge relating to its content.

To test the hypothesis that reasoning on the basis of beliefs requires effortful processing; we added a secondary task to the Experiment. Participants were initially assigned to the Secondary load condition or the Control condition. The secondary load condition required participants to randomly generate a number between 1 and 9, every second, whilst completing the reasoning questions. This secondary task is meant to heavily impact on working memory and we believe if this interferes with belief based instructions in this research, then belief based reasoning requires more effort than previously reported. The Control group had no secondary load and completed the reasoning questions alone.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below (see experiment 1).

Experiment 4: Brief, Instructions, Debrief

Brief

For the Random Number Generation Group:

The first part of this study requires you to complete a short Intelligence questionnaire which will take less than 10 minutes.

The second part to this study is interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If a person is swimming, then they are dry

Suppose a person is swimming

Is the person wet?

Or

The door is either red or blue.

The door is not blue.

Therefore, is the door red?

On some trials you will be asked to judge whether the conclusion follows logically and on other trials you will be asked whether the conclusion is believable.

Whilst making these judgments you will also be asked to say out aloud a random number between 1 and 9, every second, on the beep of a metronome. (See instructions for details). The whole procedure will take approximately 30-40 minutes.

On completion of the computer based task, you will be required to complete a short questionnaire and 3 short problems which should take between 5 – 10 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

For the Control Group the bold font section referring to the secondary task was replaced with:

"The whole procedure will take approximately 25-35 minutes".

Instructions

In the following experiment you will be presented with 128 reasoning problems of the following type:

If a child is crying, then it is sad

Suppose a child is crying

Does it follow that the child is sad?

Or

Either the snow is cold or the snow is hot

Suppose the snow is not cold

Does it follow that the snow is not hot?

For each problem you will have to answer either according to your beliefs or according to logic. When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If you finish a drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be full?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that if you finish your drink then the glass will be empty, therefore the conclusion is unbelievable.

The same applies to the following example:

Either you finish your drink and the glass is full or it is empty

Suppose you finish your drink and the glass is not full

Does it follow that the glass is empty?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is BELIEVABLE because based upon your knowledge of the world you know that if you finish your drink then the glass will be empty, therefore the conclusion is believable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentences. A valid conclusion is one that logically follows.

For example:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass is empty?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that "if you finish your drink then the glass will be full" and supposing you "finish your drink" you must logically conclude that your glass will be full. This is why the conclusion "does it follow that the glass is empty" is logically invalid.

And the same applies to the following example:

Either you finish your drink and the glass is full or it is empty

Suppose you finish your drink and the glass is not empty

Does it follow that the glass is not full?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that either the glass can be full or empty, if it is not empty then you must logically conclude that it has to be full. This is why the conclusion "does it follow that the glass is not full" is logically invalid.

For every problem you will first be presented with one statement. You have a few seconds to read the first statement before the second statement, the conclusion and response options appear on the screen. The response options will either show valid/invalid or believable/unbelievable.

If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs.

Please indicate your choice by pressing s) for VALID or BELIEVABLE and k) for INVALID or UNBELIEVABLE; depending on the available options.

During this experiment a metronome beep will sound every second (1000ms). This will be audible through your earphones. It is your task to say aloud a number between 1-9 on each beep; for the entire duration of the experiment; and your responses will be recorded on a dicta-phone. You are to continue generating random numbers whilst solving the reasoning problems.

In order to understand the concept of randomness, please consider the following:

Suppose you had written the numbers 1-9 on pieces of paper and put them in a hat. You take out one piece of paper, call out the number and return it to the hat. Then you reach for another piece of paper and do the same thing. The series of numbers you call out in that way should be random.

It is important for you to maintain a consistent response rhythm to the generation of numbers. If you should fall behind, please listen to the beep and promptly re-establish your pace.

Place the earphones on now and you will hear the metronome beep. Please practice generating random numbers ALOUD on each beep for the next 10 seconds.

Keeping your headphones on, you will now be presented with 8 practice reasoning problems, together with the metronome beep to practice the two together. These practice trials do not count towards your total performance.

If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

This concludes the practice trials. If you have any additional questions, please ask them now.

If not please PAUSE counting for 10 beeps and then say ALOUD `Starting Experiment'.

NOW COMMENCE COUNTING AND PRESS SPACE TO BEGIN WITH THE ACTUAL EXPERIMENT

The same instructions were used for the control group minus any reference to the secondary task.

Debrief

Anonymous participant ID: ________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument.

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

The aim of this research is to evaluate the degree to which logic and beliefs interfere with one another when participants are asked to make judgments based upon either the structure of an argument or knowledge relating to its content.

To test the hypothesis that reasoning on the basis of beliefs requires effortful processing; we added a secondary task to the Experiment. Participants were initially assigned to High load, or Control conditions. The High Load condition required participants to randomly generate a number between 1 and 9, every second, whilst completing the reasoning questions. This secondary task is meant to heavily impact on working memory and we believe if this interferes with belief based instructions in this research, then belief based reasoning requires more effort than previously reported. The control group had no secondary load and completed the reasoning questions alone.

With the short Intelligence test (AH4), Thinking Disposition Questionnaire (AOT) and Cognitive Reflection test (CRT), we are hoping to uncover significant relationships between cognitive ability and distinct reasoning types.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below:

Researcher:	Stephanie Howarth:
	stephanie.howarth@plymouth.ac.uk
Supervisor:	Professor Simon Handley:
	shandley@plymouth.ac.uk
Second Supervisor:	Dr Clare Walsh
	clare.walsh@plymouth.ac.uk
Alternative Contact:	Paula Simson: Secretary to the Human Ethics Committee
	paula.simson@plymouth.ac.uk
	Contact Tel: 01752 584503

References:

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences,* 7, 454-459

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25-42.

Handley, S.J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37,28-43.

Heim, A. (1970). AH4 group test of general intelligence manual. *London:Nelson*.

Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences*, *17*(2), 115-127.

Miyake, A.,Friedman, N.P., Emerson, M.J., Witzki, A.H., & Howeter, A. & Wager, T.D. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive psychology, 41*(1), 49-100.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*(3), 225-247.

Experiment 5: Brief, Instructions, Debrief

Brief

For the <u>Random Number Generation</u> Group:

The first part of this study requires you to complete 3 short problems which will take a couple of minutes.

The second part to this study is interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If a person is swimming, then they are dry

Suppose a person is swimming

Is the person wet?

Or

The door is either red or blue.

The door is not blue.

Therefore, is the door red?

On some trials you will be asked to judge whether the conclusion follows logically and on other trials you will be asked whether the conclusion is believable.

Whilst making these judgments you will also be asked to say out aloud a random number between 1 and 9, every second, to the tick of a metronome. (See instructions for details). The whole procedure will take approximately 20-25 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

For the <u>Control</u> Group the bold font section referring to the secondary task was replaced with:

"The whole procedure will take approximately 15-20 minutes".

Instructions

In the following experiment you will be presented with 128 reasoning problems of the following type:

If a child is crying, then it is sad

Suppose a child is crying

Does it follow that the child is sad?

or

Either the snow is cold or the snow is hot

Suppose the snow is not cold

Does it follow that the snow is not hot?

For each problem you will have to answer either according to your beliefs or according to logic. When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If you finish a drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass will be full?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that if you finish your drink then the glass will be empty, therefore the conclusion is unbelievable.

The same applies to the following example:

Either the snow is cold or it is hot

Suppose the snow is not hot

Does it follow that snow is cold?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is BELIEVABLE because based upon your knowledge of the world you know that snow is cold, therefore the conclusion is believable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentences. A valid conclusion is one that logically follows.

For example:

If you finish your drink then the glass will be full

Suppose you finish your drink

Does it follow that the glass is empty?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid. In this particular example the correct answer according to logic is INVALID, because the first premise states that "if you finish your drink then the glass will be full" and supposing you "finish your drink" you must logically conclude that your glass will be full. This is why the conclusion "does it follow that the glass is empty" is logically invalid.

And the same applies to the following example:

Either the snow is cold or it is hot

Suppose the snow is not hot

Does it follow that the snow is not cold?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that either the snow is cold or it is hot, if it is not hot then you must logically conclude that it cold. This is why the conclusion "does it follow that the snow is not cold" is logically invalid.

For every problem you will first be presented with one statement. You have a few seconds to read the first statement before the second statement, the conclusion and response options appear on the screen. The response options will either show valid/invalid or believable/unbelievable.

If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs.

Please indicate your choice by pressing s) for VALID or BELIEVABLE and k) for INVALID or UNBELIEVABLE; depending on the available options.

During this experiment a metronome tick will sound every second (1000)ms. This will be audible through your earphones.

It is your task to say ALOUD a number between 1-9 on each tick; for the entire duration of the experiment; and your responses will be recorded on a dicta-phone.

You will be asked to generate random numbers continuously for 5 minutes without answering any questions. Then you will be presented with a practice trial where you will be required to generate random numbers whilst solving the problems and this is what you will continue to do in the main part of the experiment. In order to understand the concept of randomness, please consider the following:

Suppose you had written the numbers 1-9 on pieces of paper and put them in a hat. You take out one piece of paper, call out the number and return it to the hat. Then you reach for another piece of paper and do the same thing. The series of numbers you call out in that way should be random.

It is important for you to maintain a consistent response rhythm to the generation of numbers. If you should fall behind, please listen to the tick and promptly re-establish your pace.

Place the earphones on now and you will hear the metronome tick.

In a moment I would like you to start generating random numbers ALOUD on each tick for the next 5 minutes. When you press the `space' bar a white screen will appear and you can start generating the numbers until further instructions.

Blank screen for 5 minutes

Keeping your headphones on, you will now be presented with 8 practice reasoning problems, together with the metronome tick to practice the two together. These practice trials do not count towards your total performance.

If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

This concludes the practice trials, if you have any additional questions, please ask them now, if not please PAUSE counting for 10 ticks and then say ALOUD `Starting Experiment':

NOW COMMENCE COUNTING AND PRESS SPACE TO BEGIN WITH THE ACTUAL EXPERIMENT.

For the Control Group, the instructions were the same apart from they were only instructed to generate random numbers for the first 5 minutes of the experiment and then instructed to stop and continue with the reasoning questions only.

Debrief

Anonymous participant ID: ________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument.

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

The aim of this research is to evaluate the degree to which logic and beliefs interfere with one another when participants are asked to make judgments based upon either the structure of an argument or knowledge relating to its content.

To test the hypothesis that reasoning on the basis of beliefs requires effortful processing; we added a secondary task to the Experiment. Participants were initially assigned to High load, or Control conditions. The High Load condition required participants to randomly generate a number between 1 and 9, every second, whilst completing the reasoning questions. This secondary task is meant to heavily impact on working memory and we believe if this interferes with belief based instructions in this research, then belief based reasoning requires more effort than previously reported. The control group had no secondary load and completed the reasoning questions alone.

With the Cognitive Reflection test (CRT), we are hoping to uncover significant relationships between cognitive ability and distinct reasoning types.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below (see experiment 1).

Experiment 7: Brief, Instructions, Debrief

Brief

Brief for the Experimental and Control condition in Experiment 7 and the Experimental condition in Experiments 7 & 8:

In this experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is RED?

Or

Either polar bears are white or they are purple

Suppose polar bears are white

Does it follow that polar bears are purple?

You will be asked to judge whether the conclusion is believable, whether it follows logically or determine whether the characteristic of the last word in the current trial is the `same' or `different' to the previous trial.

The whole procedure will take approximately 30-35 minutes.

On completion of the reasoning task, you will be required to complete another 2 Working Memory tasks and a short cognitive ability task. This should take between 15 and 20 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

Instructions

Instructions for the Experimental and Control condition in Experiment 7 and the Experimental condition in Experiments 8:

In the following experiment you will be presented with 192 reasoning problems with a break in between.

The problems will be presented as follows:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is

red?

or

Either tomatoes are red or they are white

Suppose tomatoes are white

Does it follow that tomatoes are not

red?

The last word of each problem will be in a different font.

For each problem you will have to answer either according to your beliefs, according to logic or according to the characteristics of the last word in the conclusion.

When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is

red?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is BELIEVABLE because based upon your knowledge of the world you know that cherries are red, therefore the conclusion is believable.

The same applies to the following example:

Either tomatoes are red or they are white

Suppose tomatoes are white

Does it follow that tomatoes are not

red?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that tomatoes are red, therefore the conclusion is unbelievable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentence.

A valid conclusion is one that logically follows.

For example:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is

red?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that "If the fruit is a cherry, then it is yellow" and supposing "...the fruit is a cherry" you must logically conclude that it has to be `yellow'. This is why the conclusion "does it follow that the fruit is red?" is logically invalid.

And the same applies to the following example:

Either tomatoes are red or they are white

Suppose tomatoes are white

Does it follow that tomatoes are

red?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that either the tomatoes are red or white, if they are white then you must logically conclude that they are not red. This is why the conclusion "Does it follow that tomatoes are red?" is logically invalid.

On some of the problems you will be instructed to identify whether the characteristic of the last word in the present trial matches the characteristic of the last word in the previous trial. In these cases it is irrelevant whether the statement is valid/invalid or believable/unbelievable.

You will not know whether you will be required to answer according to beliefs, logic or characteristics until the response options are presented. Therefore, it is essential that you keep track of the characteristic of the last word in each trial, throughout the experiment.

In the first slide the answer was `believable' and the last word `red' was in Raive font. Therefore the correct answer for the second slide was S) Same, since the last word was in the same font.

It is NOT important for you to know the name of the font, just indicate whether you think the font of the last word in the present trial matches the last word in the previous trial.

You only do this when the trial indicates; `Same' or `Different' as the response options.

Remember, you will not know how you should answer the trial until the response options are presented, therefore, always keep in mind the characteristic of last word from the previous trial.

In the first slide the answer was `valid' and the last word `red' was in Lucida Font. Therefore the correct answer for the second slide was K) Different, because the last word was Curlz Font, not Lucida Font.

For every problem you will first be presented with one statement. You have a few seconds to read the first statement before the second statement, the conclusion and response options appear on the screen.

The response options will either show valid/invalid, believable/unbelievable or same/different. If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs. If they show same/different indicate whether the characteristic of the last word in the present trial is the same or different from the last word of the previous trial.

When answering on the basis of belief or logic, the characteristic of the last word is irrelevant.

When responding on the basis of belief, logic or characteristic the following applies:

s) for VALID, BELIEVABLE or SAME and k) for INVALID, UNBELIEVABLE or DIFFERENT; depending on the available options.

You will now be presented with 12 practice reasoning problems. These practice trials do not count towards your total performance.

If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

That completes the practice trials. If you have any questions please ask the experimenter now. If you are ready to continue PRESS SPACE TO START THE ACTUAL EXPERIMENT.

Debrief

The following debrief was used in Experiment 7 & 8.

Anonymous participant ID: ___________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

Our research up to this point has suggested that whilst belief and logic based judgments employ the Working Memory, they may be pulling on distinct executive resources.

The aim of this research is to investigate the executive process of *memory updating* and determine whether belief and logic based reasoning command this specific executive resource.

To investigate this, we used a unique version of the N Back task. This task requires you to remember the characteristic of the last word 1 trial back and continuously update the information in your memory with each new trial.

This study should help us determine whether belief or logic based reasoning correlates with the executive resource of *updating*. We hypothesise that if logic based judgments require memory updating then the N back task should have a larger effect on logic judgments than belief based judgments.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below (see experiment 1).

Instructions for Short Stroop Task

Adapted from Raz, Shapiro, Fan & Posner (2002) and used in Experiments 7 – 9.

In this experiment you will be presented with a succession of words written in one of four of the following ink colours: red, blue, green yellow.

Before a word is presented you must focus your eyes on a fixation cross at the centre of the screen, until a word appears replacing the cross and you respond.

Your task is to identify the ink colour in which a word is written by pressing one of four keys on the keyboard.

The following 4 keys; V, B, N, and M, correspond with the colours red, blue, green, and yellow.

These are marked on your keyboard with the first letter that represents the colour.

Please use two fingers of each hand to press these response keys. For example: left middle finger for V and right index finger for N.

Please respond as quickly and accurately as possible.

There will be 3 practice trials, with feedback, to familiarise yourself with the experiment. If you have any queries once you have completed the practice trials, please speak to the experimenter.

Press next to continue.

After the practice trials:

This concludes the practice trials, if you have any question before you start the main part of the experiment, please ask the experimenter now.

If you are ready to continue on to the experimental trials, press next.

Instructions for Operation Span Task

Adapted from Turner & Engle (1989) and used in Experiments 7 – 9.

In this task you will be presented a simple equation followed by a single word to remember.

For example:

$$(3 \times 4) + 11 = 22$$

Bear

You will have to determine whether the sum of the equation given is true or false and also memorise the word presented after the equation.

If you think the answer to the sum is correct then click `Yes' if you think the answer to the sum is incorrect then click `no'.

You have 8 seconds to verify the equations so you need to respond as quickly and as accurately as possible.

After verifying the sum of the equation, you will be instructed to recall the words in the order in which they were presented.

To begin with you will be presented with 3 practice trials, to familiarise yourself with the experiment. If you have any queries once you have completed the practice trials, please speak to the experimenter.

After the practice trials:

This concludes the practice trials, if you have any question before you start the main part of the experiment, please ask the experimenter now.

If you are ready to continue on to the experimental trials, press next.

Experiment 8: Brief, Instructions

Brief

Brief for the Control group:

In this experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If the fruit is a cherry, then it is YELLOW

Suppose the fruit is a cherry

Does it follow that the fruit is RED?

Or

Either polar bears are white or they are purple

Suppose polar bears are white

Does it follow that polar bears are purple?

You will be asked to judge whether the conclusion is believable, whether it follows logically or determine whether the characteristic of the last word in the conclusion is the `same' or `different' to the last word in the first premise (sentence).

The whole procedure will take approximately 30-35 minutes.

On completion of the reasoning task, you will be required to complete another 2 Working Memory tasks and a short cognitive ability task. This should take between 15 and 20 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

Instructions

The body of information is the same as experiment 8 with specific instructions for the <u>Control</u> group outlined below:

…On some of the problems you will be instructed to identify whether the characteristic of the last word in the conclusion matches the characteristic of the last word in the first premise (sentence).

The last word in the first premise and conclusion will either be in: Lucida Font, Algerian Font, Curlz Font, Raive Font or Bradley Hand Font.

When matching the word characteristics, it is irrelevant whether the statement is valid/invalid or believable/unbelievable.

You will not know whether you will be required to answer according to beliefs, logic or characteristics until the response options are presented.

In the first slide the conclusion was `believable'. The font style of the last word in the conclusion was irrelevant.

In the second slide you would have been required to answer according to characteristics. The last word of the first premise `yellow' was in a different font style to the last word in the conclusion. So the correct response would have been: k) Different.

It is NOT important for you to know the name of the font, just indicate whether you think the font of the last word in the conclusion is the same or different to the last word in the first premise.

You only do this when the trial indicates; `Same' or `Different' as the response options.

Remember, you will not know how you should answer the trial until the response options are presented.

In the first slide the conclusion was `valid', the characteristics of the last word in the conclusion was irrelevant.

In the second slide, the last word in the first premise and the last word in the conclusion shared the same characteristics so the correct answer would have been: s)Same …

Experiment 9: Brief, Instructions, Debrief

Brief

In this experiment we are interested in examining how people respond to different instructions on reasoning tasks. You will be presented with problems of the following kind:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is red?

Or

Either polar bears are white or they are purple

Suppose polar bears are white

Does it follow that polar bears are purple?

You will be asked to judge whether the conclusion is believable, whether it follows logically or to determine the ink colour of the last word.

The whole procedure will take approximately 30-35 minutes.

On completion of the reasoning task, you will be required to complete another 2 Working Memory tasks and a short cognitive ability task. This should take between 15 and 20 minutes.

You should also be aware that you will have the right to withdraw at any point during the experiment without incurring any penalty. Your data will be identified by a code number, rather than a name and it will remain anonymous, and will not be distributed to other researchers.

Instructions

Presented to both the Experimental and Control conditions:

In the following experiment you will be presented with 192 reasoning problems with a break in between.

The problems will be presented as follows:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is

red?

or

Either tomatoes are red or they are white

Suppose tomatoes are white

Does it follow that tomatoes are not

red?

The last word of each problem will be in colour.

For each problem you will have to answer either according to your beliefs, according to logic or you will have to identify the ink colour of the last word.

When instructed to answer according to your beliefs you must answer in relation to your knowledge of what is true in the world.

For example:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is

red?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is BELIEVABLE because based upon your knowledge of the world you know that cherries are red, therefore the conclusion is believable.

The same applies to the following example:

Either tomatoes are red or they are white

Suppose tomatoes are white

Does it follow that tomatoes are not

red?

s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because based upon your knowledge of the world you know that tomatoes are red, therefore the conclusion is unbelievable.

On some of the problems you will be instructed to answer according to logic. In this case you must assume each statement is true (even if in reality it is not true) and indicate whether the conclusion follows validly from the preceding sentence.

A valid conclusion is one that logically follows.

For example:

If the fruit is a cherry, then it is yellow

Suppose the fruit is a cherry

Does it follow that the fruit is

red?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that "If the fruit is a cherry, then it is yellow" and supposing "...the fruit is a cherry" you must logically conclude that it has to be `yellow'. This is why the conclusion "does it follow that the fruit is red?" is logically invalid.

And the same applies to the following example:

Either tomatoes are red or they are white

Suppose tomatoes are white

Does it follow that tomatoes are

red?

s) VALID k) INVALID

If you think the conclusion follows logically, choose valid. If you do not think the conclusion follows logically, choose invalid.

In this particular example the correct answer according to logic is INVALID, because the first premise states that either the tomatoes are red or white, if they are white then you must logically conclude that they are not red. This is why the conclusion "Does it follow that tomatoes are red?" is logically invalid.

On some of the problems you will be instructed to identify the ink colour of the last word. In these cases it is irrelevant whether the statement is valid/invalid or believable/unbelievable.

<Example>

In the example the last word was written in red ink so the correct answer was; S) - Red

<Example>

In the last example the last word was written in white ink so the correct answer was; K) - White

For every problem you will first be presented with one statement. You have a few seconds to read the first statement before the second statement, the conclusion and response options appear on the screen. The response options will either show valid/invalid, believable/unbelievable or the choice of to 2 colours.

If they show valid/invalid you should respond on the basis of logic. If they show believable/unbelievable you should respond on the basis of your beliefs. If they show 2 colour options you should respond based on the ink colour of the last word.

When answering on the basis of belief or logic, the colour of the last word is irrelevant.

When responding on the basis of belief or logic the following applies: s) for VALID or BELIEVABLE and k) for INVALID or UNBELIEVABLE; depending on the available options.

You will now be presented with 12 practice reasoning problems. These practice trials do not count towards your total performance. If you have not familiarised yourself with the task after these practice trials, please ask the experimenter for more instructions before you start with the actual experiment.

This concludes the practice trials. If you have any questions please ask the experimenter now. If you are ready PRESS SPACE TO START THE ACTUAL EXPERIMENT.

Debrief

Anonymous participant ID: ________________________

Thank you for participating in this study.

Over the past thirty years research on thinking and reasoning has demonstrated that beliefs often lead to poor judgments, that is they interfere with our ability to reason logically and have a tendency to `lead us astray' when drawing inferences. These findings paint a rather pessimistic view of human rationality, suggesting that beliefs are often immune to change through logical argument

Recent research, however, suggests a much more optimistic view, showing that under certain circumstances, judgments relating to strongly held beliefs can themselves be undermined if they are inconsistent with the logical structure of an accompanying argument. Furthermore, reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in order to generate a novel response.

Our research up to this point has suggested that whilst belief and logic based judgments employ the Working Memory, they may be pulling on distinct executive resources.

The aim of this research is to investigate the executive process of *inhibition* and determine whether belief and logic based reasoning command this specific executive resource.

To investigate this, we used a unique version of the Stroop task. This task requires you to name the font colour of a coloured word that is either congruent or incongruent to the actual font colour. Our prepotent response is to say the colour that is written rather than the ink colour; consequently, the task requires the inhibition of this automatic inclination to read the word on incongruent trials.

This study should help us determine whether belief or logic based reasoning correlates with our ability to inhibit prepotent responses. We hypothesise that belief based judgment require the inhibition of logical interference in order to successfully complete, therefore, those who score highly on the Stroop task should be better at belief based judgments.

We would like to thank you again for taking part in our study. Should you have any further questions, or decide that you wish to withdraw your data from the study (within two weeks of taking part in the study) the contact details of the lead researcher are shown below: (see experiment 1)

Appendix B: Materials

Experiments 1 – 6

Modus Ponens stimulus (Experiments 1 – 6)

Practice Trials Stimuli

Premise 1	Premise 2	Conclusion
If hair is cut, then it is longer than before	Suppose Harry's hair is cut	Does it follow that Harry's hair is shorter than before?
If sugar is added to tea then it will taste bitter	Suppose sugar is added to tea	Does it follow that the tea will taste sweet?
If a plant is watered, then it will continue growing	Suppose a plant is watered	Does it follow that the plant stops growing?
If a person exercises, then their heart rate increases	Suppose Donna exercises	Does it follow that Donna's heart rate increases?

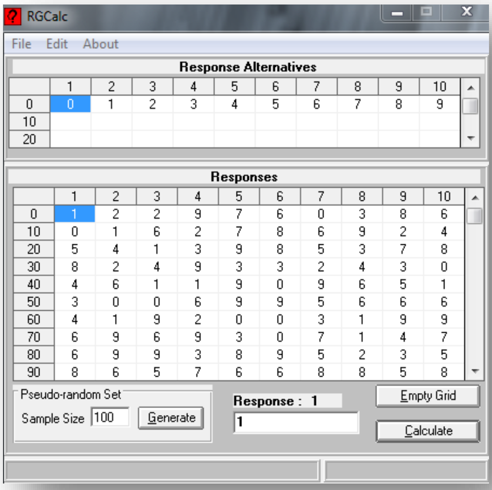
Experimental Trials Stimuli

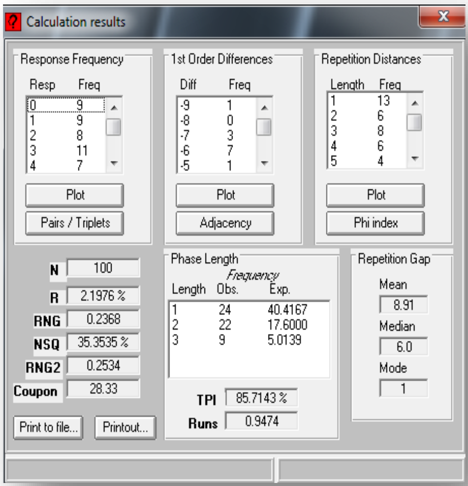
Premise 1	Premise 2	Conclusion
If a child is crying, then it is happy	Suppose a child is crying	Does it follow that the child is sad?
If a child is crying, then it is happy	Suppose a child is crying	Does it follow that the child is happy?
If a child is crying, then it is sad	Suppose a child is crying	Does it follow that the child is happy?
If a child is crying, then it is sad	Suppose a child is crying	Does it follow that the child is sad?
If a dog is barking, then it is silent	Suppose a dog is barking	Does it follow that the dog is silent?
If a dog is barking, then it is silent	Suppose a dog is barking	Does it follow that the dog is loud?
If a dog is barking, then it is loud	Suppose a dog is barking	Does it follow that the dog is loud?
If a dog is barking, then it is loud	Suppose a dog is barking	Does it follow that the dog is silent?
If a person is swimming, then he is dry	Suppose a person is swimming	Does it follow that the person is wet?
If a person is swimming, then he is dry	Suppose a person is swimming	Does it follow that the person is dry?
If a person is swimming, then he is wet	Suppose a person is swimming	Does it follow that the person is dry?
If a person is swimming, then he is wet	Suppose a person is swimming	Does it follow that the person is wet?
If a tree is old, then it is small	Suppose a tree is old	Does it follow that the tree is small?
If a tree is old, then it is small	Suppose a tree is old	Does it follow that the tree is large?
If a tree is old, then it is large	Suppose a tree is old	Does it follow that the tree

		is large?
If a tree is old, then it is large	Suppose a tree is old	Does it follow that the tree is small?
If a new computer is high-end, then it is cheap	Suppose a new computer is high-end	Does it follow that the new computer is expensive?
If a new computer is high-end, then it is cheap	Suppose a new computer is high-end	Does it follow that the new computer is cheap?
If a computer is high-end, then it is expensive	Suppose a new computer is high-end	Does it follow that the new computer is expensive?
If a computer is high-end, then it is expensive	Suppose a new computer is high-end	Does it follow that the new computer is cheap?
If a person eats too much, then they are skinny	Suppose a person eats too much	Does it follow that the person is skinny?
If a person eats too much, then they are skinny	Suppose a person eats too much	Does it follow that the person is fat?
If a person eats too much, then they are fat	Suppose a person eats too much	Does it follow that the person is skinny?
If a person eats too much, then they are fat	Suppose a person eats too much	Does it follow that the person is fat?
If the sky is grey, then it is sunny	Suppose the sky is grey	Does it follow that it is cloudy?
If the sky is grey, then it is sunny	Suppose the sky is grey	Does it follow that it is sunny?
If the sky is grey, then it is cloudy	Suppose the sky is grey	Does it follow that it is sunny?
If the sky is grey, then it is cloudy	Suppose the sky is grey	Does it follow that it is cloudy?
If the light switch is on, then it is dark inside	Suppose the light switch is on	Does it follow that it is dark inside?
If the light switch is on, then it is dark inside	Suppose the light switch is on	Does it follow that it is bright inside?
If the light switch is on, then it is bright inside	Suppose the light switch is on	Does it follow that it is bright inside?
If the light switch is on, then it is bright inside	Suppose the light switch is on	Does it follow that it is dark inside?
If a hamster gets fed, then it will die	Suppose a hamster gets fed	Does it follow that the hamster will live?
If a hamster gets fed, then it will die	Suppose a hamster gets fed	Does it follow that the hamster will die?
If a hamster gets fed, then it will live	Suppose a hamster gets fed	Does it follow that the hamster will live?
If a hamster gets fed, then it will live	Suppose a hamster gets fed	Does it follow that the hamster will die?
If an aeroplane runs out of fuel, then it will fly	Suppose an aeroplane runs out of fuel	Does it follow that the aeroplane will fly?
If an aeroplane runs out of fuel, then it will fly	Suppose an aeroplane runs out of fuel	Does it follow that the aeroplane will crash?
If an aeroplane runs out of fuel, then it will crash	Suppose an aeroplane runs out of fuel	Does it follow that the aeroplane will fly?
If an aeroplane runs out of fuel, then it will crash	Suppose an aeroplane runs out of fuel	Does it follow that the aeroplane will crash?
If your mother bakes cookies,	Suppose your mother	Does it follow that you will

you will be sad	bakes cookies	be happy?
If your mother bakes cookies, you will be sad	Suppose your mother bakes cookies	Does it follow that you will be sad?
If your mother bakes cookies, you will be happy	Suppose your mother bakes cookies	Does it follow that you will be sad?
If your mother bakes cookies, you will be happy	Suppose your mother bakes cookies	Does it follow that you will be happy?
If a bear catches a fish, it is slow	Suppose a bear catches a fish	Does it follow that the bear is slow?
If a bear catches a fish, it is slow	Suppose a bear catches a fish	Does it follow that the bear is quick?
If a bear catches a fish, it is quick	Suppose a bear catches a fish	Does it follow that the bear is quick?
If a bear catches a fish, it is quick	Suppose a bear catches a fish	Does it follow that the bear is slow?
If a toilet has been flushed, then it is dirty	Suppose a toilet is flushed	Does it follow that the toilet is clean?
If a toilet has been flushed, then it is dirty	Suppose a toilet is flushed	Does it follow that the toilet is dirty?
If a toilet has been flushed, then it is clean	Suppose a toilet is flushed	Does it follow that the toilet is clean?
If a toilet has been flushed, then it is clean	Suppose a toilet is flushed	Does it follow that the toilet is dirty?
If a glass falls on the floor then it is intact	Suppose a glass falls on the floor	Does it follow that the glass is intact?
If a glass falls on the floor then it is intact	Suppose a glass falls on the floor	Does it follow that the glass is broken?
If a glass falls on the floor then it is broken	Suppose a glass falls on the floor	Does it follow that the glass is intact?
If a glass falls on the floor then it is broken	Suppose a glass falls on the floor	Does it follow that the glass is broken?
If heat bill has not been paid in time then it is warm inside	Suppose the heat bill has not been paid in time	Does it follow that it is cold inside?
If heat bill has not been paid in time then it is warm inside	Suppose the heat bill has not been paid in time	Does it follow that it is warm inside?
If the heat bill has not been paid in time then it is cold inside	Suppose the heat bill has not been paid in time	Does it follow that it is warm inside?
If heat bill has not been paid in time then it is cold inside	Suppose the heat bill has not been paid in time	Does it follow that it is cold inside?
If a singer has a sore throat, his singing will be nice	Suppose the singer has a sore throat	Does it follow that his singing will be nice?
If a singer has a sore throat, his singing will be nice	Suppose the singer has a sore throat	Does it follow that his singing will be bad?
If a singer has a sore throat, his singing will be bad	Suppose the singer has a sore throat	Does it follow that his singing will be bad?
If a singer has a sore throat, his singing will be bad	Suppose the singer has a sore throat	Does it follow that his singing will be nice?

RGCalc program (Experiments 2 – 5)





Disjunctive stimulus (Experiments 4 – 6)

Practice Trials Stimuli

Premise 1	Premise 2	Conclusion	Item Type
Either ants are insects or they are arachnids	Suppose an ant is not an insect	Does it follow that the ant is an arachnid?	SDA
Either pubs sell drinks or they sell books	Suppose a pub sells books	Does it follow that the pub does not sell drinks?	SDB
Either photographs are visual or they are auditory	Suppose a photograph is not auditory	Does it follow that the photograph is visual?	SDA
Either vegetarians eat vegetables or they eat meat	Suppose a vegetarian eats vegetables	Does it follow that the vegetarian eats meat?	SDB

Experimental Trials Stimuli

Premise 1	Premise 2	Conclusion	Item Type
Either cats are mammals or they are plants	Suppose a cat is not a mammal	Does it follow that the cat is not a plant?	SDA
Either cats are mammals or they are plants	Suppose a cat is not a mammal	Does it follow that the cat is a plant?	SDA
Either cats are mammals or they are plants	Suppose a cat is not a plant	Does it follow that the cat is not a mammal?	SDA
Either cats are mammals or they are plants	Suppose a cat is not a plant	Does it follow that the cat is a mammal?	SDA
Either the sky is blue or it is green	Suppose the sky is not blue	Does it follow that the sky is green?	SDA
Either the sky is blue or it is green.	Suppose the sky is not blue	Does it follow that the sky is not green?	SDA
Either the sky is blue or it is green	Suppose the sky is not green	Does it follow that the sky is blue?	SDA
Either the sky is blue or it is green	Suppose the sky is not green	Does it follow that the sky is not blue?	SDA
Either the sun is yellow or it is blue	Suppose the sun is blue	Does it follow that the sun is yellow?	SDB
Either the sun is yellow or it is blue	Suppose the sun is blue	Does it follow that the sun is not yellow?	SDB
Either the sun is yellow or it is blue	Suppose the sun is yellow	Does it follow that the sun is blue?	SDB
Either the sun is yellow or it is blue	Suppose the sun is yellow	Does it follow that the sun is not blue?	SDB
Either the sea is blue or it is pink	Suppose the sea is pink	Does it follow that the sea is not blue?	SDB
Either the sea is blue or it is pink	Suppose the sea is pink	Does it follow that the sea is blue?	SDB
Either the sea is blue or it is pink	Suppose the sea is blue	Does it follow that the sea is not pink?	SDB
Either the sea is blue or it is pink	Suppose the sea is blue	Does it follow that the sea is pink?	SDB
Either parrots can fly or they	Suppose a parrot	Does it follow that the	SDA

can swim	cannot fly	parrot cannot swim?	
Either parrots can fly or they can swim	Suppose a parrot cannot fly	Does it follow that the parrot can swim?	SDA
Either parrots can fly or they can swim	Suppose a parrot cannot swim	Does it follow that the parrot cannot fly?	SDA
Either parrots can fly or they can swim	Suppose a parrot cannot swim	Does it follow that the parrot can fly?	SDA
Either obese people are fat or they are skinny	Suppose an obese person is not fat	Does it follow that the obese person is skinny?	SDA
Either obese people are fat or they are skinny	Suppose an obese person is not fat	Does it follow that the obese person is not skinny?	SDA
Either obese people are fat or they are skinny	Suppose an obese person is not skinny	Does it follow that the obese person is fat?	SDA
Either obese people are fat or they are skinny	Suppose an obese person is not skinny	Does it follow that the obese person is not fat?	SDA
Either skyscrapers are huge or they are tiny	Suppose a skyscraper is not huge	Does it follow that the skyscraper is tiny?	SDB
Either skyscrapers are huge or they are tiny	Suppose a skyscraper is not huge	Does it follow that the skyscraper is not tiny?	SDB
Either skyscrapers are huge or they are tiny	Suppose a skyscraper is not tiny	Does it follow that the skyscraper is not huge?	SDB
Either skyscrapers are huge or they are tiny	Suppose a skyscraper is not tiny	Does it follow that the skyscraper is huge?	SDB
Either roses are flowers or they are machines	Suppose a rose is a machine	Does it follow that the rose is not a flower?	SDB
Either roses are flowers or they are machines	Suppose a rose is a machine	Does it follow that the rose is a flower?	SDB
Either roses are flowers or they are machines	Suppose a rose is a flower	Does it follow that the rose is not a machine?	SDB
Either roses are flowers or they are machines	Suppose a rose is a flower	Does it follow that the rose is a machine?	SDB
Either sentences are made out of words or they are made out of bricks	Suppose a sentence is not made out of words	Does it follow that the sentence is not made out of bricks?	SDA
Either sentences are made out of words or they are made out of bricks	Suppose a sentence is not made out of words	Does it follow that the sentence is made out of bricks?	SDA
Either sentences are made out of words or they are made out of bricks	Suppose a sentence is not made out of bricks	Does it follow that the sentence is not made out of words?	SDA
Either sentences are made out of words or they are made out of bricks	Suppose a sentence is not made out of bricks	Does it follow that the sentence is made out of words?	SDA
Either mice eat cheese or they eat steel	Suppose a mouse does not eat cheese	Does it follow that the mouse eats steel?	SDA
Either mice eat cheese or they eat steel	Suppose a mouse does not eat cheese	Does it follow that the mouse does not eat steel?	SDA
Either mice eat cheese or they eat steel	Suppose a mouse does not eat steel	Does it follow that the mouse eats cheese?	SDA
Either mice eat cheese or they eat steel	Suppose a mouse does not eat steel	Does it follow that the mouse does not eat cheese?	SDA

Either alligators eat meat or they eat plastic	Suppose an alligator eats plastic	Does it follow that the alligator does not eat meat?	SDB
Either alligators eat meat or they eat plastic	Suppose an alligator eats plastic	Does it follow that the alligator eats meat?	SDB
Either alligators eat meat or they eat plastic	Suppose an alligator eats meat	Does it follow that the alligator does not eat plastic?	SDB
Either alligators eat meat or they eat plastic	Suppose an alligator eats meat	Does it follow that the alligator eats plastic?	SDB
Either circles are round or they are square	Suppose a circle is square	Does it follow that the circle is round?	SDB
Either circles are round or they are square	Suppose a circle is square	Does it follow that the circle not round?	SDB
Either circles are round or they are square	Suppose a circle is round	Does it follow that the circle is square?	SDB
Either circles are round or they are square	Suppose a circle is round	Does it follow that the circle is not square?	SDB
Either monkeys are primates or they are rodents	Suppose a monkey is not a primate	Does it follow that the monkey is not a rodent?	SDA
Either monkeys are primates or they are rodents	Suppose a monkey is not a primate	Does it follow that the monkey is a rodent?	SDA
Either monkeys are primates or they are rodents	Suppose a monkey is not a rodent	Does it follow that the monkey is not a primate?	SDA
Either monkeys are primates or they are rodents	Suppose a monkey is not a rodent	Does it follow that the monkey is a primate?	SDA
Either flowers are organic or they are vehicles	Suppose a flower is not organic	Does it follow that the flower is a vehicle?	SDA
Either flowers are organic or they are vehicles	Suppose a flower is not organic	Does it follow that the flower is not a vehicle?	SDA
Either flowers are organic or they are vehicles	Suppose a flower is not a vehicle	Does it follow that the flower is organic?	SDA
Either flowers are organic or they are vehicles	Suppose a flower is not a vehicle	Does it follow that the flower is not organic?	SDA
Either shovels are tools or they are stationary	Suppose a shovel is stationary	Does it follow that the shovel is not a tool?	SDB
Either shovels are tools or they are stationary	Suppose a shovel is stationary	Does it follow that the shovel is a tool?	SDB
Either shovels are tools or they are stationary	Suppose a shovel is a tool	Does it follow that the shovel is stationary?	SDB
Either shovels are tools or they are stationary	Suppose a shovel is a tool	Does it follow that the shovel is not stationary?	SDB
Either spears are weapons or they are thermostats	Suppose a spear is a thermostat	Does it follow that the spear is a weapon?	SDB
Either spears are weapons or they are thermostats	Suppose a spear is a thermostat	Does it follow that the spear is not a weapon?	SDB
Either spears are weapons or they are thermostats	Suppose a spear is a weapon	Does it follow that the spear is not a thermostat?	SDB
Either spears are weapons or they are thermostats	Suppose a spear is a weapon	Does it follow that the spear is a thermostat?	SDB

Item type refers to the category of disjunctives used. SDA are disjunctives of the following type: A or B, not A, therefore B. Negation of the first part which makes the second part true. SDB are of the type: A or B, A, therefore not B. The first part is true which makes the second part not true. These are valid because they are **exclusive disjunctives**, meaning A and B can never be true at the same time. An equal number of the SDA and SDB were used in Experiments 4 & 5.

Individual Differences stimulus (Experiments 4)

Questions from the AH4 Group Test of General Intelligence Part 1 (Heim, 1970).

Q1	1, 2, 3, 4, 5, 6, 7, 8, 9. Multiply the middle one of these figures by 2.
Q2	*1 2 3 4 5* *Easy* means the opposite of.... problem, simple, difficult, always, cannot.
Q3	15, 35, 55, 75, 95... What number comes next?
Q4	*1 2 3 4 5* *Seed* is to *plant* as *egg* is to ... tree, bird, pollen, oats, potato.
Q5	Here are three figures: 234. Divide the biggest figure by the smallest and add the results to the figure printed immediately after the smallest figure.
Q6	*1 2 3 4 5* Rich means the same as ... poor, wealthy, high, new, lucky.
Q7	1, 2, 3, 4, 5, 6, 7, 8, 9. Write down the fourth figure to the left of 7.
Q8	*1 2 3 4 5* Right means the opposite of ... action, good, careless, wrong, motive.
Q9	1, 2, 4, 8, 16... what number comes next?
Q10	*1 2 3 4 5* *Foot* is to *leg* as *hand* is to ... body, finger, tail, limb, arm.
Q11	Here are three figures: 327. Subtract the smallest figure from the biggest and multiply the result by the figure printed immediately before the biggest figure.
Q12	*1 2 3 4 5* *Old* means the same as... decaying, tired, aged, youth, mended.
Q13	1, 2, 3, 4, 5, 6, 7, 8, 9. Add the first five figures together and subtract them from the sum of the last four.
Q14	*1 2 3 4 5* *Lost* means the opposite of... winning, draw, found, alone, mislaid.
Q15	3, 3, 7, 7, 11... What number comes next?
Q16	*1 2 3 4 5* *Army* is to *navy* as *soldier* is to ... airman, sea, service, sailor, uniform.
Q17	

	Here are three figures: 132. Divide the biggest figure by the smallest and add the result to the figure printed immediately after the smallest figure.
Q18	*1 2 3 4 5* *Portion* means the same as… some, whole, part, any, cake.
Q19	If a castle is bigger than a cottage, write down the second of these figures: 1, 2, 3, 4, 5, 6, 7, 8, 9. If it is not, write down the sixth.
Q20	*1 2 3 4 5* *Up* means the opposite of… short, small, low, down, young.
Q21	$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}$… what number comes next?
Q22	*1 2 3 4 5* *Seeing* is to *picture* as *hearing* is to… sight, sculpture, ear, song, deaf.
Q23	Here are three figures: 189. Subtract the smallest figure from the biggest and multiply the result by the figure printed immediately before the biggest figure.
Q24	*1 2 3 4 5* *Ill* means the same as… health, fever, dirty, mumps, sick.
Q25	Write down the number of the letters in the fourth word of this sentence.
Q26	*1 2 3 4 5* *Near* means the opposite of… close, road, speed, far, distance.
Q27	2, 3, 5, 8 , 12… What number comes next?
Q28	*1 2 3 4 5* Legs are to running as teeth are to … chattering, walking, eating, biting, arms.
Q29	Here are three figures: 627. Add the largest two figures together and divide the total by the smallest figure.
Q30	*1 2 3 4 5* *Scarce* means the same as… unobtainable, lack, unique, rare, frightened.
Q31	If Z is the last letter of the alphabet and is B does not come before A, write down the fifth of these figures: 1, 2, 3, 4, 5, 6, 7, 8, 9. Otherwise, write down the last one.
Q32	*1 2 3 4 5* *Never* means the opposite of… rarely, always, now, will, forget.
Q33	1, 2, 4, 5, 7 … What number comes next?
Q34	*1 2 3 4 5* *Sky* is to *ground* as *ceiling* is to… roof, down, floor, rug, high.
Q35	Here are three figures: 823. Divide the biggest figure by the smallest and add the result to the figure printed immediately after the smallest figure.
Q36	*1 2 3 4 5* *Odd* means the same as… strange, even, one, man, number.
Q37	If 8 is more than 3, write down 7, unless 3 is more than 7, in which case write 8.
Q38	*1 2 3 4 5* *War* means the opposite of… suffering, joy, dictatorship, inflation, peace.
Q39	11, 12, 10, 13, 9… What number comes next?

Q40	*1 2 3 4 5* *When* is to *where* as *time* is to... how, why, space, length, relativity.
Q41	Here is a row of figures: 1, 2, 3, 4, 5, 6, 7, 8, 9. Write down the figure from this row which, when added to another number smaller than it, would make 17.
Q42	*1 2 3 4 5* *Backwards* means the same as... upside-down, reversed, stop, forward, gear.
Q43	If 20 is more than 3 time 5, write down the figure 2, unless 14 is less than 16, in which case write 7.
Q44	*1 2 3 4 5* *Multiplication* is the opposite of... subtraction, addition, mathematics, figures, division.
Q45	0.9, 1.1, 1.3, 1.5, 1.7... What number comes next?
Q46	*1 2 3 4 5* *Autumn* is to *Winter* as *October* is to... April, July, Spring, rain, January.
Q47	Here are three figures: 456. Subtract the smallest figure from the biggest and multiply the result by the figure printed immediately before the biggest figure.
Q48	*1 2 3 4 5* *Prevent* means the same as... avoid, cure, allow, deter, help.
Q49	Write down the total number of letters contained in the words in this sentence.
Q50	*1 2 3 4 5* *Permanent* means the opposite of... part-time, ever, changing, temporary, stable.
Q51	100, 81, 64, 49, 36... What number comes next?
Q52	*1 2 3 4 5* *Fact* is to *fiction* as *historian* is to... history, book, novelist, teacher, story.
Q53	Here are three figures: 934. Divide the biggest figure by the smallest and add the result to the figure printed immediately after the smallest figure.
Q54	*1 2 3 4 5* *Industrious* means the same as... busy, hard-working, energetic, overworked, happy.
Q55	If G is the seventh letter of the alphabet and Wednesday is not a month of the year, divide 63 by 7. Otherwise subtract 3 from 5. Write down your answer.
Q56	*1 2 3 4 5* *Dangerous* means the opposite of... brave, cowardly, situation, safe, bravado.
Q57	0.1, 1.3, 2.5, 3.7, 4.9... What number comes next?
Q58	*1 2 3 4 5* *Motive* is to *method* as *why* is to... wherefore, reason, how, because, where.
Q59	Here are three figures: 847. Divide the biggest figure by the smallest and add the result to the figure printed immediately after the smallest figure.
Q60	*1 2 3 4 5* *Flat* means the same as... straight, level, uneven, oblique, inclined.
Q61	0, 2, 8, 26, 80... What number comes next?
Q62	

	Doubt means the opposite of... wonder, certainty, correct, dubious, indefinite.
Q63	130, 118, 107, 97, 98... What number comes next?
Q64	*The day after tomorrow* is to *the day before yesterday* as *Wednesday* is to... *1* *2* *3* *4* *5* Friday, Saturday, Sunday, Monday, Tuesday.
Q65	Here are three figures: 948. Divide the biggest figure by the smallest and add the result to the figure printed immediately after the smallest figure.

Statements presented in the Actively Open-minded Thinking Scale questionnaire (McPherson & Stanovich, 2007; Stanovich & West, 2007).

"Please circle the number that corresponds to the alternative that best describes your opinion. There are no right or wrong answers so do not spend too much time deciding on an answer. The first thing that comes to mind is probably the best response. Be sure the number on the answer sheet corresponds to the number of the statement to which you are responding. There is no time limit, but work as quickly as possible".

1. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

2. What beliefs you hold have more to do with your own personal character than the experiences that may have given rise to them

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

3. I tend to classify people as either for me or against me.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

4. A person should always consider new possibilities.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

5. There are two kinds of people in this world: those who are for the truth and those who are against the truth.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

6. Changing your mind is a sign of weakness.

1	2	3	4	5	6
Disagree Strongly	Disagree Moderately	Disagree Slightly	Agree Slightly	Agree Moderately	Agree Strongly

7. I believe we should look to our religious authorities for decisions on moral issues.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

8. I think there are many wrong ways, but only one right way, to almost anything.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

9. It makes me happy and proud when someone famous holds the same beliefs that I do.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

10. Difficulties can usually be overcome by thinking about the problem, rather than through waiting for good fortune.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

11. There are a number of people I have come to hate because of the things they stand for.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

12. Abandoning a previous belief is a sign of strong character.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

13. No one can talk me out of something I know is right.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

14. Basically, I know everything I need to know about the important things in life.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

15. It is important to persevere in your beliefs even when evidence is brought to bear against them.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

16. Considering too many different opinions often leads to bad decisions.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

17. There are basically two kinds of people in this world, good and bad.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

18. I consider myself broad-minded and tolerant of other people's lifestyles.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

19. Certain beliefs are just too important to abandon no matter how good a case can be made against them.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

20. Most people just don't know what's good for them.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

21. It is a noble thing when someone holds the same beliefs as their parents.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

22. Coming to decisions quickly is a sign of wisdom.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

23. I believe that loyalty to one's ideals and principles is more important than "open-mindedness."

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

24. Of all the different philosophies which exist in the world there is probably only one which is correct.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

25. My beliefs would not have been very different if I had been raised by a different set of parents.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

26. If I think longer about a problem I will be more likely to solve it.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

27. I believe that the different ideas of right and wrong that people in other societies have may be valid for them.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

28. Even if my environment (family, neighbourhood, schools) had been different, I probably would have the same religious views.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

29. There is nothing wrong with being undecided about many issues.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

30. I believe that laws and social policies should change to reflect the needs of a changing world.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

31. My blood boils over whenever a person stubbornly refuses to admit he's wrong.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

32. I believe that the "new morality" of permissiveness is no morality at all.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

33. One should disregard evidence that conflicts with your established beliefs.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

34. Someone who attacks my beliefs is not insulting me personally.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

35. A group which tolerates too much difference of opinion among its members cannot exist for long.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

36. Often, when people criticize me, they don't have their facts straight.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

37. Beliefs should always be revised in response to new information or evidence.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

38. I think that if people don't know what they believe in by the time they're 25, there's something wrong with them.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

39. I believe letting students hear controversial speakers can only confuse and mislead them.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

40. Intuition is the best guide in making decisions.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

41. People should always take into consideration evidence that goes against their beliefs.

1___________2___________3___________4___________5___________6

Disagree Strongly Disagree Moderately Disagree Slightly Agree Slightly Agree Moderately Agree Strongly

The Cognitive Reflection Task (Frederick, 2005)

Please answer these questions

1. A bat and a ball cost £1.10 in total. The bat costs £1 more than the ball. How much does the ball cost? ______________

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ______________

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ______________

Experiments 7 – 9

Modus Ponens stimulus (Experiments 7 – 9)

Practice Trials Stimuli

Premise 1	Premise 2	Conclusion
If the fruit is a cherry, then it is yellow	Suppose the fruit is a cherry	Does it follow that the fruit is red?
If the vegetable is a courgette, then it is green	Suppose vegetable is a courgette	Does it follow that the vegetable is pink?
If the drink is Guinness, then it is red	Suppose the drink is Guinness	Does it follow that the drink is black?
If the swan is a cygnet, then it is grey	Suppose the swan is a cygnet	Does it follow that the swan is grey?

Experimental Trials Stimuli

Premise 1	Premise 2	Conclusion
If the car has stopped, then the traffic lights are green	Suppose the car has stopped	Does it follow that the traffic lights are green?
If a child is crying, it feels blue	Suppose a child is crying	Does it follow that the child feels pink?
If the clouds are stormy, then they are red	Suppose the clouds are stormy	Does it follow that the clouds are grey?
If the fruit is a strawberry, then it is red	Suppose the fruit is a strawberry	Does it follow that the fruit is red?
If the teeth are cleaned, then they will be brown	Suppose the teeth are cleaned	Does it follow that the teeth will be white?
If it is night time, then the sky is black	Suppose it is night time	Does it follow that the sky is black?
If the banana is ripe, then it is red	Suppose the banana is ripe	Does it follow that the banana is red?
If the bird is a canary, then it is yellow	Suppose the bird is a canary	Does it follow that the bird is blue?
If the gemstone is an Emerald, then it is yellow	Suppose the gemstone is an Emerald	Does it follow that the gemstone is yellow?
If the plant is a Shamrock, then it is green	Suppose the plant is a Shamrock	Does it follow that the plant is pink?
If a bird is a dove, then it is orange	Suppose a bird is a dove	Does it follow that the bird is white?
If the drink is Pepsi, then the drink is brown	Suppose the drink is Pepsi	Does it follow that the drink is brown?
If the vegetable is a carrot, then it is blue	Suppose the vegetable is a carrot	Does it follow that it is orange?
If the vegetable is a beetroot, then it is purple	Suppose the vegetable is a beetroot	Does it follow that the vegetable is purple?
If the herb is basil, then it is blue	Suppose the herb is basil	Does it follow that the herb is blue?
If the flower is a Daffodil, then it	Suppose the flower is a	Does it follow that the flower

is yellow	Daffodil	is purple?
If a child is crying, then it feels pink	Suppose a child is crying	Does it follow that the child feels pink?
If the car has stopped, then the traffic lights are red	Suppose the car has stopped	Does it follow that the traffic lights are green?
If the fruit is a strawberry, then it is purple	Suppose the fruit is a strawberry	Does it follow that the fruit is red?
If the clouds are stormy, then they are grey	Suppose the clouds are stormy	Does its follow that the clouds are grey?
If it is night time, then the sky is yellow	Suppose it is night time	Does it follow that the sky is black?
If the teeth are cleaned, then they will be white	Suppose the teeth are cleaned	Does it follow that the teeth will be white?
If the bird is a canary, then it is blue	Suppose the bird is a canary	Does it follow that the bird is blue?
If the banana is ripe, then it is yellow	Suppose the banana is ripe	Does it follow that the banana is red?
If the plant is a Shamrock, then it is pink	Suppose the plant is a Shamrock	Does it follow that the plant is pink?
If the gemstone is an Emerald, then it is green	Suppose the gemstone is an Emerald	Does it follow that the gemstone is yellow?
If the drink is Pepsi, then the drink is blue	Suppose the drink is Pepsi	Does it follow that the drink is brown?
If a bird is a dove, then it is white	Suppose a bird is a dove	Does it follow that the bird is white?
If the vegetable is a beetroot, then it is white	Suppose the vegetable is a beetroot	Does it follow that the vegetable is purple?
If the vegetable is carrot, then it is orange	Suppose the vegetable is a carrot	Does it follow that it is orange?
If the flower is a Daffodil, then it is purple	Suppose the flower is a Daffodil	Does it follow that the flower is purple?
If the herb is basil, then it is green	Suppose the herb is basil	Does it follow that the herb is blue?
If a child is crying, then it feels pink	Suppose a child is crying	Does it follow that the child feels blue?
If the car has stopped, then the traffic lights are red	Suppose the car has stopped	Does it follow that the traffic lights are red?
If the fruit is a strawberry, then it is purple	Suppose the fruit is a strawberry	Does it follow that the fruit is purple?
If the clouds are stormy, then they are grey	Suppose the clouds are stormy	Does it follow that the clouds are red?
If it is night time, then the sky is yellow	Suppose it is night time	Does it follow that the sky is yellow?
If the teeth are cleaned, then they will be white	Suppose the teeth are cleaned	Does it follow that the teeth will be brown?
If the bird is a canary, then it is blue	Suppose the bird is a canary	Does it follow that the bird is yellow?
If the banana is ripe, then it is yellow	Suppose the banana is ripe	Does it follow that the banana is yellow?
If the plant is a Shamrock, then it is pink	Suppose the plant is a Shamrock	Does it follow that the plant is green?
If the gemstone is an Emerald,	Suppose the gemstone	Does it follow that the

then it is green	is an Emerald	gemstone is green?
If the drink is Pepsi, then the drink is blue	Suppose the drink is Pepsi	Does it follow that the drink is blue?
If a bird is a dove, then it is white	Suppose a bird is a dove	Does it follow that the bird is orange?
If the vegetable is a beetroot, then it is white	Suppose the vegetable is a beetroot	Does it follow that the vegetable is white?
If the vegetable is carrot, then it is orange	Suppose the vegetable is a carrot	Does it follow that it is blue?
If the flower is a Daffodil, then it is purple	Suppose the flower is a Daffodil	Does it follow that the flower is yellow?
If the herb is basil, then it is green	Suppose the herb is basil	Does it follow that the herb is green?
If the car has stopped, then the traffic lights are green	Suppose the car has stopped	Does it follow that the traffic lights are red?
If a child is crying, then it feels blue	Suppose a child is crying	Does it follow that the child feels blue?
If the clouds are stormy, then they are red	Suppose the clouds are stormy	Does it follow that the clouds are red?
If the fruit is a strawberry, then it is red	Suppose the fruit is a strawberry	Does it follow that the fruit is purple?
If the teeth are cleaned, then they will be brown	Suppose the teeth are cleaned	Does it follow that the teeth will be brown?
If it is night time, then the sky is black	Suppose it is night time	Does it follow that the sky is yellow?
If the banana is ripe, then it is red	Suppose the banana is ripe	Does it follow that the banana is yellow?
If the bird is a canary, then it is yellow	Suppose the bird is a canary	Does it follow that the bird is yellow?
If the gemstone is an Emerald, then it is yellow	Suppose the gemstone is an Emerald	Does it follow that the gemstone is green?
If the plant is a Shamrock, then it is green	Suppose the plant is a Shamrock	Does it follow that the plant is green?
If a bird is a dove, then it is orange	Suppose a bird is a dove	Does it follow that the bird is orange?
If the drink is Pepsi, then the drink is brown	Suppose the drink is Pepsi	Does it follow that the drink is blue?
If the vegetable is a carrot, then it is blue	Suppose the vegetable is a carrot	Does it follow that it is blue?
If the vegetable is a beetroot, then it is purple	Suppose the vegetable is a beetroot	Does it follow that the vegetable is white?
If the herb is basil, then it is blue	Suppose the herb is basil	Does it follow that the herb is green?
If the flower is a Daffodil, then it is yellow	Suppose the flower is a Daffodil	Does it follow that the flower is yellow?

Disjunctive stimulus (Experiments 7 – 9)

Practice Trials Stimuli

Premise 1	**Premise 2**	**Conclusion**	**Item Type**
Either grass is green or it is blue	Suppose grass is not blue	Does it follow that grass is green?	SDA
Either tomatoes are red or they are white	Suppose tomatoes are white	Does it follow that tomatoes are not red?	SDB
Either milk is white or it is purple	Suppose milk is not white	Does it follow that milk is purple?	SDA
Either coffee is black or it is green	Suppose coffee is not green	Does it follow that coffee is not black?	SDB

Experimental Trials Stimuli

Premise 1	**Premise 2**	**Conclusion**	**Item Type**
Either lemons are yellow or they are purple	Suppose lemons are not yellow	Does it follow that lemons are purple?	SDA
Either snow is white or it is orange	Suppose snow is not orange	Does it follow that snow is not white?	SDA
Either blood is red or it is white	Suppose blood is not red	Does it follow that blood is not white?	SDA
Either the sky is blue or it is green	Suppose the sky is not green	Does it follow that the sky is blue?	SDA
Either cucumbers are green or they are blue	Suppose cucumbers are not green	Does it follow that cucumbers are blue?	SDA
Either elephants are grey or red	Suppose elephants are not red	Does it follow that elephants are not grey?	SDA
Either sapphires are blue or they are yellow	Suppose sapphires are not blue	Does it follow that sapphires are not yellow?	SDA
Either charcoal is black or it is green	Suppose charcoal is not green	Does it follow that charcoal is black?	SDA
Either snow is white or it is orange	Suppose snow is not white	Does it follow that snow is orange?	SDA
Either lemons are yellow or they are purple	Suppose lemons are not purple	Does it follow that lemons are not yellow?	SDA
Either the sky is blue or it is green.	Suppose the sky is not blue	Does it follow that the sky is not green?	SDA
Either blood is red or it is white	Suppose blood is not white	Does it follow that blood is red?	SDA
Either elephants are grey or red	Suppose elephants are not grey	Does it follow that elephants are red?	SDA
Either cucumbers are green or they are blue	Suppose cucumbers are not blue	Does it follow that cucumbers are not green?	SDA
Either charcoal is black or it is green	Suppose charcoal is not black	Does it follow that charcoal is not green?	SDA
Either sapphires are blue or they are yellow	Suppose sapphires are not yellow	Does it follow that sapphires are blue?	SDA
Either swans are white or	Suppose swans are	Does it follow that swans	SDB

black	black	are white?	
Either the sun is yellow or it is blue	Suppose the sun is yellow	Does it follow that the sun is blue?	SDB
Either flamingos are pink or purple	Suppose flamingos are purple	Does it follow that flamingos are not pink?	SDB
Either the sea is blue or it is pink	Suppose the sea is blue	Does it follow that the sea is not pink?	SDB
Either Ravens are black or yellow	Suppose Ravens are black	Does it follow that Ravens are yellow?	SDB
Either Ravens are black or yellow	Suppose Ravens are yellow	Does it follow that Ravens are black?	SDB
Either plums are purple or green	Suppose plums are green	Does it follow that plums are purple?	SDB
Either plums are purple or green	Suppose plums are purple	Does it follow that plums are not green?	SDB
Either the sun is yellow or it is blue	Suppose the sun is blue	Does it follow that the sun is yellow?	SDB
Either swans are white or black	Suppose swans are white	Does it follow that swans are black?	SDB
Either the sea is blue or it is pink	Suppose the sea is pink	Does it follow that the sea is not blue?	SDB
Either flamingos are pink or purple	Suppose flamingos are pink	Does it follow that flamingos are not purple?	SDB
Either custard is yellow or it is black	Suppose custard is black	Does it follow that custard is yellow?	SDB
Either custard is yellow or it is black	Suppose custard is yellow	Does it follow that custard is not black?	SDB
Either ketchup is red or brown	Suppose ketchup is brown	Does it follow that ketchup is not red?	SDB
Either ketchup is red or brown	Suppose ketchup is red	Does it follow that ketchup is brown?	SDB
Either snow is white or it is orange	Suppose snow is not white	Does it follow that snow is not orange?	SDA
Either lemons are yellow or they are purple	Suppose lemons are not purple	Does it follow that lemons are yellow?	SDA
Either the sky is blue or it is green	Suppose the sky is not blue	Does it follow that the sky is green?	SDA
Either blood is red or it is white	Suppose blood is not white	Does it follow that blood is not red?	SDA
Either swans are white or they are black	Suppose swans are black	Does it follow that swans are not white?	SDB
Either swans are white or they are black	Suppose swans are white	Does it follow swans are not black?	SDB
Either flamingos are pink or purple	Suppose flamingos are purple	Does it follow that flamingos are pink?	SDB
Either flamingos are pink or purple	Suppose flamingos are pink	Does it follow that flamingos are purple?	SDB
Either cucumbers are green or they are blue	Suppose cucumbers are not green	Does it follow that cucumbers are not blue?	SDA
Either elephants are grey or they are red	Suppose elephants are not red	Does it follow that elephants are grey?	SDA
Either sapphires are blue	Suppose sapphires	Does it follow that	SDA

or they are yellow	are not blue	sapphires are yellow?	
Either charcoal is black or it is green	Suppose charcoal is not green	Does it follow that charcoal is not black?	SDA
Either custard is yellow or it is black	Suppose custard is black	Does it follow that custard is not yellow?	SDB
Either Ravens are black or they are yellow	Suppose Ravens are yellow	Does it follow that Ravens are not black?	SDB
Either ketchup is red or brown	Suppose ketchup is brown	Does it follow that ketchup is red?	SDB
Either plums are purple or they are green	Suppose plums are purple	Does it follow that plums are green?	SDB
Either lemons are yellow or they are purple	Suppose lemons are not yellow	Does it follow that lemons are not purple?	SDA
Either snow is white or it is orange	Suppose snow is not orange	Does it follow that snow is white?	SDA
Either blood is red or it is white	Suppose blood is not red	Does it follow that blood is white?	SDA
Either the sky is blue or it is green	Suppose the sky is not green	Does it follow that the sky is not blue?	SDA
Either the sun is yellow or it is blue	Suppose the sun is blue	Does it follow that the sun is yellow?	SDB
Either the sun is yellow or it is blue	Suppose the sun is yellow	Does it follow that the sun is blue?	SDB
Either the sea is blue or it is pink	Suppose the sea is pink	Does it follow that the sea is blue?	SDB
Either the sea is blue or it is pink	Suppose the sea is blue	Does it follow that the sea is pink?	SDB
Either cucumbers are green or they are blue	Suppose cucumbers are not blue	Does it follow that cucumbers are green?	SDA
Either elephants are grey or they are red	Suppose elephants are not grey	Does it follow that elephants are not red?	SDA
Either sapphires are blue or they are yellow	Suppose sapphires are not yellow	Does it follow that sapphires are not blue?	SDA
Either charcoal is black or it is green	Suppose charcoal is not black	Does it follow that charcoal is green?	SDA
Either custard is yellow or it is black	Suppose custard is yellow	Does it follow that custard is black?	SDB
Either Ravens are black or they are yellow	Suppose Ravens are black	Does it follow that Ravens are yellow?	SDB
Either ketchup is red or brown	Suppose ketchup is red	Does it follow that ketchup is not brown?	SDB
Either plums are purple or they are green	Suppose plums are green	Does it follow that plums are not purple?	SDB

Individual Differences stimulus (Experiments 7 – 9)

Short Stroop Task Stimulus

Word	Ink Colour	Condition
Blue	Red	Incongruent
Blue	Green	Incongruent
Blue	Yellow	Incongruent
Red	Blue	Incongruent
Red	Green	Incongruent
Red	Yellow	Incongruent
Yellow	Red	Incongruent
Yellow	Blue	Incongruent
Yellow	Green	Incongruent
Green	Blue	Incongruent
Green	Red	Incongruent
Green	Yellow	Incongruent
Flower	Red	Neutral
Flower	Green	Neutral
Flower	Yellow	Neutral
Ship	Blue	Neutral
Ship	Red	Neutral
Ship	Green	Neutral
Lot	Yellow	Neutral
Lot	Blue	Neutral
Lot	Red	Neutral
Knife	Green	Neutral
Knife	Yellow	Neutral
Knife	Blue	Neutral
Blue	Blue	Congruent
Blue	Blue	Congruent
Blue	Blue	Congruent
Red	Red	Congruent
Red	Red	Congruent
Red	Red	Congruent
Yellow	Yellow	Congruent
Yellow	Yellow	Congruent
Yellow	Yellow	Congruent
Green	Green	Congruent
Green	Green	Congruent
Green	Green	Congruent

Operation Span Stimulus

Number in set	Equation	Word
2	(10 ÷ 2) - 3 = 2 ?	SEA
2	(10 ÷ 10) - 1 = 2 ?	CLASS
2	(17 ÷ 1) + 2 = 7 ?	PAINT
2	(3 ÷ 1) - 2 = 3 ?	CLOUD
2	(2 x 1) - 1 = 1 ?	PIPE
2	(10 ÷ 1) + 3 = 13 ?	EAR
2	(9 x 2) + 1 = 18 ?	FLAME
2	(9 ÷ 1) - 7 = 4 ?	BIKE
2	(8 x 4) - 2 = 32 ?	BEAN
2	(9 x 3) - 3 =24 ?	ARM
2	(4 ÷ 1) + 1 = 4 ?	GROUND
2	(10 ÷ 1) - 1 = 9 ?	HOLE
3	(8 x 4) + 2 = 34 ?	DAD
3	(6 x 3) + 2 = 17 ?	KID
3	(6 ÷ 3) + 2 = 5 ?	FORK
3	(6 x 2) - 3 = 10 ?	JAIL
3	(8 ÷ 2) + 4 = 2 ?	HAT
3	(8 ÷ 2) - 1 = 3 ?	LAMP
3	(9 ÷ 1) - 5 = 4 ?	CAVE
3	(6 ÷ 2) - 2 = 2 ?	BACK
3	(7 x 2) - 1 = 14 ?	HALL
4	(6 x 2) - 2 = 10 ?	FERN
4	(2 x 2) + 1 =4 ?	MAN
4	(7 x 1) + 6 = 13 ?	WORLD
4	(3 ÷ 1) + 3 = 6 ?	DRILL
4	(10 ÷ 1) + 1 = 10 ?	CALF
4	(4 x 4) + 1 = 17 ?	FISH
4	(3 x 3) -1 = 8 ?	CHEEK
4	(3 x 1) + 2 = 2 ?	BREAD
4	(4 ÷ 2) + 1 = 6 ?	GERM
4	(5 ÷ 5) + 1 = 2 ?	DOCK
4	(2 x 3) + 1 = 4 ?	GAME
4	(9 ÷ 3) - 2 = 1 ?	NERVE
5	(10 ÷ 2) - 4 = 3 ?	WAX
5	(5 ÷ 1) + 4 = 9 ?	TIN
5	(10 x 2) + 3 = 23 ?	CHURCH
5	(7 ÷ 1) + 6 = 12 ?	BEACH
5	(3 x 2) + 1 = 6 ?	CARD
5	(6 x 4) + 1 = 25	JOB
5	(9 ÷ 3) - 1 = 2 ?	CONE
5	(8 ÷ 1) - 6 = 4 ?	BRASS
5	(9 x 1) + 9 = 1 ?	STREET
5	(4 x 6) - 4 = 20 ?	SHIN
5	(2 x 9) + 7 = 24 ?	ROAD
5	(5 ÷ 5) + 5 = 5 ?	HUMAN
5	(6 x 7) + 4 = 54?	POST
5	(3 x 9) + 1 = 28 ?	WASH
5	(9 x 9) + 2 = 82 ?	CAMP

The Cognitive Reflection Task (Experiments 7 – 9)

Please answer these questions

1. A bat and a ball cost £1.10 in total. The bat costs £1 more than the ball. How much does the ball cost? ______________

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ______________

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ______________

Appendix C: Statistical Tables

The following analyses are presented in the order in which they are presented in the thesis.

Experiment 1

Accuracy

Table C1.1 Repeated measures ANOVA on excluded data – N = 75

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.173	1	.173	3.435	.068	.045
Instruction x Load (high or low load)	.001	1	.001	.020	.889	.000
Error (Instruction)	3.685	73	.050			
Problem Type (Conflict/No-conflict)	3.175	1	3.175	51.057	.000	.412
Problem Type x Load	.001	1	.001	.017	.897	.000
Error (Problem Type)	4.540	73	.062			
Instruction x Problem Type	.140	1	.140	3.717	.058	.048
Instruction x Problem Type x Load	.058	1	.058	1.547	.218	.021
Error (Instruction * Problem Type)	2.758	73	.038			

Note: Load is a between subjects factor

Table C1.2 Between subject effects on excluded data – N = 75

	SS	df	MS	F	*p*	η_p^2
Intercept	536.419	1	536.419	3272.210	.000	.987
Load	.181	1	.181	1.102	.297	.015
Error	11.967	73	.164			

Table C1.3 Repeated measures ANOVA on full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.131	1	.131	2.189	.143	.027
Instruction x Load (high or low load)	.025	1	.025	.412	.523	.005
Error (Instruction)	4.727	79	.060			
Problem Type (Conflict/No-conflict)	4.750	1	4.750	59.022	.000	.428
Problem Type x Load	.006	1	.006	.076	.783	.001
Error (Problem Type)	6.358	79	.080			
Instruction x Problem Type	.101	1	.101	1.468	.229	.018
Instruction x Problem Type x Load	8.883E-5	1	8.883E-5	.001	.971	.000
Error (Instruction * Problem Type)	5.411	79	.068			

Note: Load is a between subjects factor

Table C1.4 Between subject effects on full data set–N = 81

	SS	df	MS	F	p	η_p^2
Intercept	554.108	1	554.108	2739.552	.000	.972
Load	.122	1	.122	.602	.440	.008
Error	15.979	79	.202			

<u>Latencies</u>

Table C1.5 Repeated measures ANOVA on excluded data – N = 75

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	7469937.847	1	7469937.847	36.921	.000	.336
Instruction x Load (high or low load)	777079.282	1	777079.282	3.841	.054	.050
Error (Instruction)	147696033.58	73	202323.337			
Problem Type (Conflict/No-conflict)	10269660.24	1	10269660.24	36.928	.000	.336
Problem Type x Load	531888.539	1	531888.539	1.913	.171	.026
Error (Problem Type)	20301246.08	73	278099.261			
Instruction x Problem Type	529269.195	1	529269.195	1.780	.186	.024
Instruction x Problem Type x Load	436182.799	1	436182.799	1.467	.230	.020
Error (Instruction * Problem Type)	21706064.53	73	297343.350			

Note: Load is a between subjects factor

Table C1.6 Between subject effects on excluded data – N = 75

	SS	df	MS	F	p	η_p^2
Intercept	4.082E9	1	4.082E9	2028.485	.000	.965
Load	10713.367	1	10713.367	.005	.942	.000
Error	1.469E8	73	2012351.110			

Table C1.7 Repeated measures ANOVA on full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	7387475.292	1	7387475.292	23.257	.000	.227
Instruction x Load (high or low load)	777665.560	1	777665.560	2.448	.122	.030
Error (Instruction)	25094394.06	79	317650.558			
Problem Type (Conflict/No-conflict)	15216636.35	1	15216636.35	26.994	.000	.255
Problem Type x Load	2002904.124	1	2002904.124	3.553	.063	.043
Error (Problem Type)	44533218.63	79	563711.628			
Instruction x Problem Type	407499.502	1	407499.502	.730	.395	.009
Instruction x Problem Type x Load	331080.763	1	331080.763	.593	.444	.007
Error (Instruction * Problem Type)	44096495.89	79	297343.350			

Note: Load is a between subjects factor

Table C1.8 Between subject effects on full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Intercept	4.670E9	1	4.670E9	1422.926	.000	.947
Load	157392.388	1	157392.388	.048	.827	.001
Error	2.4593E8	79	3281895.575			

Experiment 2

Accuracy

Table C2.1 Repeated measures ANOVA on excluded data – N = 67

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.034	1	.034	.973	.328	.015
Instruction x Condition (RNG,AS or Control)	.397	2	.198	5.675	.005	.151
Error (Instruction)	2.236	64	.035			
Problem Type (Conflict/No-conflict)	2.739	1	2.739	50.555	.000	.441
Problem Type x Condition	.064	2	.032	.588	.559	.018
Error (Problem Type)	3.467	64	.054			
Instruction x Problem Type	.044	1	.044	.697	.407	.011
Instruction x Problem Type x Condition	.003	2	.001	.022	.978	.001
Error (Instruction * Problem Type)	4.060	64	.063			

Note: Condition is a between subjects factor

Table C2.2 Between subject effects on excluded data – N = 67

	SS	df	MS	F	*p*	η_p^2
Intercept	343.993	1	343.993	1896.683	.000	.967
Condition	6.201	2	3.101	17.097	.000	.348
Error	11.607	64	.181			

Table C2.3 Post Hoc Analysis (LSD) of Condition

	AS	Control	RNG
AS		.001	.018
Control	.001		.000
RNG	.018	.000	

Table C2.4 Repeated measures ANOVA on full data set – N = 74

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	166.824	1	166.824	1.101	.298	.015
Instruction x Condition (RNG,AS or Control)	769.215	2	384.607	2.538	.086	.067
Error (Instruction)	10758.577	71	151.529			
Problem Type (Conflict/No-conflict)	8654.372	1	8654.372	48.003	.000	.403
Problem Type x Condition	420.387	2	210.193	1.167	.317	.032
Error (Problem Type)	12787.093	71	180.100			
Instruction x Problem Type	340.063	1	340.063	1.462	.231	.020
Instruction x Problem Type x Condition	218.487	2	109.243	.470	.627	.013
Error (Instruction * Problem Type)	16518.343	71	232.653			

Note: Condition is a between subjects factor

Table C2.5 Between subject effects on full data set – N = 74

	SS	df	MS	F	*p*	η_p^2
Intercept	2082412.388	1	2082412.388	4553.925	.000	.985
Condition	15713.250	2	7856.625	17.181	.000	.326
Error	32466.781	71	457.295			

Follow Up Analysis

Table C2.6 Repeated measures ANOVA comparing RNG and AS condition - excluded data set – N = 43

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.005	1	.005	.112	.740	.003
Instruction x Condition (RNG, AS)	.142	1	.142	2.932	.094	.067
Error (Instruction)	1.980	41	.048			
Problem Type (Conflict/No-conflict)	1.363	1	1.363	32.842	.000	.445
Problem Type x Condition	.076	1	.076	1.832	.183	.043
Error (Problem Type)	1.702	41	.042			
Instruction x Problem Type	.119	1	.119	1.480	.231	.035
Instruction x Problem Type x Condition	.014	1	.014	.171	.681	.004
Error (Instruction * Problem Type)	3.309	41	.081			

Note: Condition is a between subjects factor

Table C2.7 Between subject effects on RNG and AS condition - excluded data set– N = 43

	SS	df	MS	F	*p*	η_p^2
Intercept	185.235	1	185.235	816.079	.000	.952
Condition	1.062	1	1.062	4.678	.036	.102
Error	9.306	41	.227			

Table C2.8 T-Test for the difference between belief and logic instruction between the RNG and AS Condition –excluded data set – N = 43

	Mean	Std.Dv.	N	Diff.	*t*	*df*	*p*
Belief AS	.833807	.1405461	22				
Belief RNG	.788690	.1397210	21	.0451163	1.055	41	.297
Logic AS	.860795	.1230228	22				
Logic RNG	.776786	.1235406	21	.0840097	2.234	41	.031

Table C2.9 Repeated measures ANOVA comparing RNG and Control Condition - excluded data set– N = 45

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.018	1	.018	.430	.515	.010
Instruction x Condition (RNG, Control)	.197	1	.197	4.690	.036	.098
Error (Instruction)	1.802	43	.042			
Problem Type (Conflict/No-conflict)	1.490	1	1.490	28.332	.000	.397
Problem Type x Condition	.096	1	.096	1.829	.183	.041
Error (Problem Type)	2.261	43	.053			
Instruction x Problem Type	.093	1	.093	.970	.330	.022
Instruction x Problem Type x Condition	.028	1	.028	.297	.589	.007
Error (Instruction * Problem Type)	4.109	43	.096			

Note: Condition is a between subjects factor

Table C2.10 Between subject effects on RNG and Control Condition - excluded data set – N = 45

	SS	df	MS	F	*p*	η_p^2
Intercept	234.485	1	234.485	2003.433	.000	.979
Condition	6.097	1	6.097	52.089	.000	.548
Error	5.033	43	.117			

Table C2.11 T-Test for the difference between belief and logic instruction between the RNG and Control Condition –excluded data set – N = 46

	Mean	Std.Dv.	N	Diff.	*t*	*df*	*p*
Belief RNG	.789773	.1364482	22				
Belief Control	.917969	.0730338	24	.1281960	4.020	44	.000
Logic RNG	.769886	.1248308	22				
Logic Control	.950521	.0469127	24	.1806345	6.604	44	.000

Table C2.12 Repeated measures ANOVA comparing AS and Control Condition - excluded data set – N = 46

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.275	1	.275	8.062	.007	.155
Instruction x Condition (AS, Control)	.004	1	.004	.105	.747	.002
Error (Instruction)	1.504	44	.034			
Problem Type (Conflict/No-conflict)	2.313	1	2.313	38.827	.000	.469
Problem Type x Condition	.001	1	.001	.014	.906	.000
Error (Problem Type)	2.621	44	.060			
Instruction x Problem Type	.035	1	.035	.718	.402	.016
Instruction x Problem Type x Condition	.002	1	.002	.049	.825	.001
Error (Instruction * Problem Type)	2.137	44	.049			

Note: Condition is a between subjects factor

Table C2.13 between subject effects on AS and Control Condition - excluded data set– N = 46

	SS	df	MS	F	*p*	η_p^2
Intercept	274.465	1	274.465	1262.653	.000	.969
Condition	2.058	1	2.058	10.218	.003	.188
Error	8.862	44	.201			

Latencies

Table C2.14 Repeated measures ANOVA on excluded data set – N = 67

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	8168111.499	1	8168111.499	11.779	.001	.155
Instruction x Condition (RNG, AS & Control)	854381.711	2	427190.856	.616	.543	.019
Error (Instruction)	44379358.331	64	693427.474			
Problem Type (Conflict/No-conflict)	4924959.005	1	4924959.005	9.098	.004	.124
Problem Type x Condition	558535.320	2	279267.660	.516	.599	.016
Error (Problem Type)	34643417.290	64	541303.395			
Instruction x Problem Type	1936951.996	1	1936951.996	4.286	.042	.063
Instruction x Problem Type x Condition	11749.711	2	5874.855	.013	.987	.000
Error (Instruction * Problem Type)	28921260.489	64	451894.695			

Note: Condition is a between subjects factor

Table C2.15 Between subject effects on excluded data set – N = 67

	SS	df	MS	F	*p*	η_p^2
Intercept	3749660399.632	1	3749660399.632	1254.417	.000	.951
Condition	6863368.493	2	3431684.247	1.148	.324	.035
Error	3749660399.632	1	3749660399.632	1254.417	.000	.951

Table C2.16 Repeated measures ANOVA on full data set – N = 74

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	21240754.72	1	21240754.72	7.841	.008	.095
Instruction x Condition	1227574.593	2	613787.296	.216	.806	.006
Error (Instruction)	2.016E8	71	2839403.667			
Problem Type (Conflict/No-conflict)	12169573.30	1	1269573.30	3.689	.059	.049
Problem Type x Condition	4162313.611	2	2081156.805	.631	.535	.017
Error (Problem Type)	2.342E8	71	3298938.351			
Instruction x Problem Type	8742993.382	1	8742993.832	3.333	.072	.045
Instruction x Problem Type x Condition	5530089.635	2	2765044.818	1.054	.345	.029
Error (Instruction * Problem Type)	1.863E8	71	2623354.328			

Note: Condition is a between subjects factor

Table C2.17 Between subject effects on full data set – N = 74

	SS	df	MS	F	*p*	η_p^2
Intercept	4.540E9	1	4.540E9	685.547	.000	.906
Condition	7653676.581	2	3826838.291	.578	.564	.016
Error	4.702E8	71	6623160.400			

Experiment 3

Accuracy

Table C3.1 Repeated measures ANOVA on excluded data set – N = 58

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.242	1	1.242	14.460	.000	.205
Instruction x Condition (RNG or Control)	.002	1	.001	.002	.891	.000
Error (Instruction)	4.810	56	.086			
Problem Type (Conflict/No-conflict)	4.025	1	4.025	54.650	.000	.494
Problem Type x Condition	.102	1	.102	1.383	.245	.024
Error (Problem Type)	4.125	56	.074			
Instruction x Problem Type	.015	1	.015	.189	.666	.003
Instruction x Problem Type x Condition	.064	1	.064	.790	.378	.014
Error (Instruction * Problem Type)	4.570	56	.082			

Note: Condition is a between subjects factor

Table C3.2 Between subject effects on excluded data set – N = 58

	SS	df	MS	F	*p*	η_p^2
Intercept	335.903	1	335.903	1234.008	.000	.957
Condition	1.055	1	1.055	3.874	.054	.065
Error	15.243	56	.272			

Table C3.3 Repeated measures ANOVA on full data set – N = 76

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.297	1	.297	7.026	.010	.087
Instruction x Condition (RNG or Control)	.222	1	.022	.531	.468	.007
Error (Instruction)	3.125	74	.042			
Problem Type (Conflict/No-conflict)	1.952	1	1.952	68.596	.000	.481
Problem Type x Condition	.058	1	.058	2.026	.159	.027
Error (Problem Type)	2.106	74	.028			
Instruction x Problem Type	.093	1	.093	2.503	.118	.033
Instruction x Problem Type x Condition	.080	1	.080	2.148	.147	.028
Error (Instruction * Problem Type)	2.759	74	.037			

Note: Condition is a between subjects factor

Table BC.4 Between subject effects on full data set– N = 76

	SS	df	MS	F	*p*	η_p^2
Intercept	201.194	1	201.194	2394.757	.000	.970
Condition	.637	1	.637	7.587	.007	.093
Error	6.217	74	.084			

<u>Latencies</u>

Table C3.5 Repeated measures ANOVA on excluded data set – N = 58

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	280.784	1	280.784	.000	.992	.000
Instruction x Condition (RNG or Control)	2258888.502	1	2258888.502	.920	.341	.016
Error (Instruction)	1.374E8	56	2454021.896			
Problem Type (Conflict/No-conflict)	13770978.602	1	13770978.602	5.661	.021	.092
Problem Type x Condition	650269.747	1	650269.747	.270	.607	.005
Error (Problem Type)	1.362E8	56	2432790.847			
Instruction x Problem Type	20529075.858	1	20529075.858	8.414	.005	.131
Instruction x Problem Type x Condition	2158302.730	1	2158302.730	.885	.351	.016
Error (Instruction * Problem Type)	1.366E8	56	2439743.763			

Note: Condition is a between subjects factor

Table C3.6 Between subject effects on excluded data set – N = 58

	SS	df	MS	F	*p*	η_p^2
Intercept	1.173E10	1	1.173E10	608.792	.000	.916
Condition	3.388E8	1	3.388E8	17.579	.000	.239
Error	1.079E9	56	19274194.761			

Table C3.7 Repeated measures ANOVA on full data set – N = 76

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	4094787.801	1	4094787.801	.844	.361	.011
Instruction x Condition (RNG or Control)	1744227.569	1	1744227.569	.359	.551	.005
Error (Instruction)	3.591E8	74	4853377.682			
Problem Type (Conflict/No-conflict)	23780731.057	1	23780731.057	4.728	.033	.060
Problem Type x Condition	1489430.880	1	1489430.880	.296	.588	.004
Error (Problem Type)	3.722E8	74	5029282.851			
Instruction x Problem Type	11563204.844	1	11563204.844	1.495	.255	.020
Instruction x Problem Type x Condition	6423045.882	1	6423045.882	.831	.365	.011
Error (Instruction * Problem Type)	5.723E8	74	7733264.163			

Note: Condition is a between subjects factor

Table C3.8 Between subject effects on full data set – N = 76

	SS	df	MS	F	*p*	η_p^2
Intercept	1.518E10	1	1.518E10	415.766	.000	.849
Condition	4.296E8	1	4.296E8	11.765	.001	.137
Error	2.702E9	74	36516410.570			

<u>Follow Up Analysis</u>

Table C3.7 Showing the impact of conflict on belief judgments: excluded data set – N = 58

	SS	df	MS	F	*p*	η_p^2
Problem Type	36461424.201	1	36461424.201	14.896	.000	.210
Error	137072910.599	56	2447730.546			

Table C3.8 Showing the impact of conflict on logic judgments: excluded data set – N = 58

	SS	df	MS	F	*p*	η_p^2
Problem Type	157980.475	1	157980.475	.065	.799	.001
Error	135281384.432	56	2415739.008			

Experiment 4

Accuracy

Table C4.1 Repeated measures ANOVA on excluded data set – N = 67

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.345	1	2.345	19.266	.000	.229
Instruction x Condition (RNG or Control)	.240	1	.240	1.968	.165	.029
Error (Instruction)	7.911	65	.122			
Problem Type (Conflict/No-conflict)	6.868	1	6.868	66.225	.000	.505
Problem Type x Condition	.144	1	.144	1.391	.243	.021
Error (Problem Type)	6.741	65	.104			
Complexity (MP/Disjunctives)	3.026	1	3.026	43.964	.000	.403
Complexity x Condition	.070	1	.070	1.015	.317	.015
Error (Complexity)	4.474	65	.069			
Instruction x Problem Type	2.019	1	2.019	20.278	.000	.239
Instruction x Problem Type x Condition	.042	1	.042	.429	.515	.007
Error (Instruction * Problem Type)	6.439	65	.099			
Instruction x Complexity	.126	1	.126	3.189	.079	.047
Instruction x Complexity x Condition	.033	1	.033	.846	.361	.013
Error (Instruction * Complexity)	2.565	65	.039			
Problem x Complexity	.394	1	.394	10.962	.002	.144
Problem x Complexity x Condition	.137	1	.137	3.807	.055	.055
Error (Problem * Complexity)	2.337	65	.036			
Instruction x Problem x Complexity	.055	1	.055	1.627	.207	.024
Instruction x Problem x Complexity x Condition	.001	1	.001	.042	.839	.001
Error (Instruction x Problem x Complexity)	2.211	65	.034			

Note: Condition is a between subjects factor

Table C4.2 Between subject effects on excluded data set – N = 67

	SS	df	MS	F	*p*	η_p^2
Intercept	677.746	1	677.746	1390.461	.000	.955
Condition	6.431	1	6.431	13.193	.001	.169
Error	31.683	65	.487			

Table C4.3 Repeated measures ANOVA on full data set – N = 80

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.528	1	.528	11.601	.001	.129
Instruction x Condition (RNG or Control)	.003	1	.003	.063	.802	.001
Error (Instruction)	3.548	78	.045			
Problem Type (Conflict/No-conflict)	2.967	1	2.967	59.335	.000	.432
Problem Type x Condition	.000	1	.000	.008	.928	.000
Error (Problem Type)	3.900	78	.050			
Complexity (MP/Disjunctives)	.415	1	.415	43.476	.000	.358
Complexity x Condition	.011	1	.011	1.168	.283	.015
Error (Complexity)	.745	78	.010			
Instruction x Problem Type	.402	1	.402	10.016	.002	.114
Instruction x Problem Type x Condition	.083	1	.083	2.069	.154	.026
Error (Instruction * Problem Type)	3.130	78	.040			
Instruction x Complexity	.001	1	.001	.128	.722	.002
Instruction x Complexity x Condition	.025	1	.025	2.666	.107	.033
Error (Instruction * Complexity)	.730	78	.009			
Problem x Complexity	.053	1	.053	5.848	.018	.070
Problem x Complexity x Condition	.003	1	.003	.325	.570	.004
Error (Problem * Complexity)	.703	78	.009			
Instruction x Problem x Complexity	.014	1	.014	1.450	.232	.018
Instruction x Problem x Complexity x Condition	.000	1	.000	.019	.892	.000
Error (Instruction x Problem x Complexity)	.753	78	.010			

Note: Condition is a between subjects factor

Table C4.4 Between subject effects on full data set – N = 80

	SS	df	MS	F	*p*	η_p^2
Intercept	419.537	1	419.537	2759.145	.000	.973
Condition	1.617	1	1.617	10.636	.002	.120
Error	11.860	78	.152			

Follow Up Analysis

Table C4.5 Repeated measures ANOVA on excluded data for MP arguments – N = 67

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.779	1	1.779	23.537	.000	.266
Instruction x Condition (RNG or Control)	.047	1	.047	.622	.433	.009
Error (Instruction)	4.912	65	.076			
Problem Type (Conflict/No-conflict)	5.276	1	5.276	70.356	.000	.520
Problem Type x Condition	.281	1	.281	3.748	.057	.055
Error (Problem Type)	4.875	65	.075			
Instruction x Problem Type	.703	1	.703	9.613	.003	.129
Instruction x Problem Type x Condition	.014	1	.014	.194	.661	.003
Error (Instruction * Problem Type)	4.752	65	.073			

Note: Condition is a between subjects factor

Table C4.6 Repeated measures ANOVA on excluded data for Disjunctive arguments – N = 67

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.692	1	.692	8.085	.006	.111
Instruction x Condition (RNG or Control)	.226	1	.226	2.638	.109	.039
Error (Instruction)	5.565	65	.086			
Problem Type (Conflict/No-conflict)	1.985	1	1.985	30.703	.000	.321
Problem Type x Condition	4.766E-005	1	4.766E-005	.001	.978	.000
Error (Problem Type)	4.203	65	.065			
Instruction x Problem Type	1.371	1	1.371	22.868	.000	.260
Instruction x Problem Type x Condition	.030	1	.030	.496	.484	.008
Error (Instruction * Problem Type)	3.898	65	.060			

Note: Condition is a between subjects factor

Latencies

Table C4.7 Repeated measures ANOVA on excluded data set – N = 67

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	3905627.639	1	3905627.639	.808	.372	.012
Instruction x Condition (RNG or Control)	913642.368	1	913642.368	.189	.665	.003
Error (Instruction)	314158764.629	65	4833211.764			
Problem Type (Conflict/No-conflict)	10603267.102	1	10603267.102	2.611	.111	.039
Problem Type x Condition	530963.598	1	530963.598	.131	.719	.002
Error (Problem Type)	263948074.213	65	4060739.603			
Complexity (MP/Disjunctives)	375920794.093	1	375920794.093	66.329	.000	.505
Complexity x Condition	659676.764	1	659676.764	.116	.734	.002
Error (Complexity)	368386426.489	65	5667483.484			
Instruction x Problem Type	10352651.083	1	10352651.083	2.650	.108	.039
Instruction x Problem Type x Condition	15830.448	1	15830.448	.004	.949	.000
Error (Instruction * Problem Type)	253886692.927	65	3905949.122			
Instruction x Complexity	13311259.561	1	13311259.561	3.389	.070	.050
Instruction x Complexity x Condition	6630675.851	1	6630675.851	1.688	.198	.025
Error (Instruction * Complexity)	255340590.832	65	3928316.782			
Problem x Complexity	537434.622	1	537434.622	.181	.672	.003
Problem x Complexity x Condition	152449.980	1	152449.980	.051	.822	.001
Error (Problem * Complexity)	193490645.741	65	2976779.165			
Instruction x Problem x Complexity	183731.729	1	183731.729	.063	.803	.001
Instruction x Problem x Complexity x Condition	961552.042	1	961552.042	.329	.568	.005
Error (Instruction x Problem x Complexity)	189864508.449	65	2920992.438			

Note: Condition is a between subjects factor

Table C4.8 Between subject effects on excluded data set – N = 67

	SS	df	MS	F	*p*	η_p^2
Intercept	29026745643.943	1	29026745643.943	951.326	.000	.936
Condition	656200167.407	1	656200167.407	21.506	.000	.249
Error	1983271662.601	65	30511871.732			

Table C4.9 Repeated measures ANOVA on full data set – N = 80

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	4.975E8	1	4.975E8	42.703	.000	.354
Instruction x Condition (RNG or Control)	3099545.257	1	3099545.257	.266	.607	.003
Error (Instruction)	9.088E8	78	11650982.704			
Problem Type (Conflict/No-conflict)	28247200.137	1	28247200.137	4.351	.040	.053
Problem Type x Condition	12918860.708	1	12918860.708	1.990	.162	.025
Error (Problem Type)	5.063E8	78	6491400.830			
Complexity (MP/Disjunctives)	30743881.493	1	30743881.493	4.805	.031	.058
Complexity x Condition	28089.039	1	28089.039	.004	.947	.000
Error (Complexity)	4.991E8	78	6398811.315			
Instruction x Problem Type	5576530.499	1	5576530.499	.937	.336	.012
Instruction x Problem Type x Condition	6410843.252	1	6410843.252	1.077	.303	.014
Error (Instruction * Problem Type)	4.643E8	78	5952346.115			
Instruction x Complexity	3281172.515	1	3281172.515	.580	.449	.007
Instruction x Complexity x Condition	2220.643	1	2220.643	.000	.984	.000
Error (Instruction * Complexity)	4.414E8	78	5658893.391			
Problem x Complexity	2476763.679	1	2476763.679	.447	.506	.006
Problem x Complexity x Condition	28293280.235	1	28293280.235	5.105	.027	.061
Error (Problem * Complexity)	4.323E8	78	5542601.001			
Instruction x Problem x Complexity	11427.645	1	11427.645	.003	.954	.000
Instruction x Problem x Complexity x Condition	3073443.162	1	3073443.162	.894	.347	.011
Error (Instruction x Problem x Complexity)	2.681E8	78	3436886.299			

Note: Condition is a between subjects factor

Table C4.10 between subject effects on full data set – N = 80

	SS	df	MS	F	p	η_p^2
Intercept	3.852E10	1	3.852E10	734.088	.000	.904
Condition	1.302E9	1	1.302E9	24.811	.000	.241
Error	4.093E9	78	52479737.322			

Individual Differences

Table C4.11 Repeated measures ANOVA on excluded accuracy data set – N = 72 – AH4

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.523	1	2.523	22.286	.000	.241
Instruction x AH4 (high or low group)	.453	1	.453	4.004	.049	.054
Error (Instruction)	7.926	70	.113			
Problem Type (Conflict/No-conflict)	6.571	1	6.571	61.949	.000	.469
Problem Type x AH4	.013	1	.013	.119	.731	.002
Error (Problem Type)	7.425	70	.106			
Complexity (MP/Disjunctives)	2.874	1	2.874	42.395	.000	.377
Complexity x AH4	.009	1	.009	.130	.719	.002
Error (Complexity)	4.745	70	.068			
Instruction x Problem Type	1.817	1	1.817	19.254	.000	.216
Instruction x Problem Type x AH4	.181	1	.181	1.919	.170	.027
Error (Instruction * Problem Type)	6.605	70	.094			
Instruction x Complexity	.090	1	.090	2.316	.133	.032
Instruction x Complexity x AH4	.001	1	.001	.016	.901	.000
Error (Instruction * Complexity)	2.722	70	.039			
Problem x Complexity	.286	1	.286	7.227	.009	.094
Problem x Complexity x AH4	.002	1	.002	.059	.809	.001
Error (Problem * Complexity)	2.767	70	.040			
Instruction x Problem x Complexity	.089	1	.089	2.571	.113	.035
Instruction x Problem x Complexity x AH4	.000	1	.000	.006	.938	.000
Error (Instruction x Problem x Complexity)	2.433	70	.035			

Note: AH4 group (high and low) as a between subjects factor

Table C4.12 Between subject effects on excluded data set – N = 72 – AH4

	SS	df	MS	F	*p*	η_p^2
Intercept	725.463	1	725.463	1348.045	.000	.951
AH4 group	5.170	1	5.170	9.607	.003	.121
Error	37.671	70	.538			

Table C4.13 Repeated measures ANOVA on full accuracy data set – N = 80 – AH4

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.533	1	.533	11.827	.001	.132
Instruction x AH4 (high or low group)	.037	1	.037	.824	.367	.010
Error (Instruction)	3.514	78	.045			
Problem Type (Conflict/No-conflict)	2.965	1	2.965	59.319	.000	.432
Problem Type x AH4	.002	1	.002	.031	.861	.000
Error (Problem Type)	3.899	78	.050			
Complexity (MP/Disjunctives)	.411	1	.411	42.435	.000	.352
Complexity x AH4	.001	1	.001	.102	.750	.001
Error (Complexity)	.755	78	.010			
Instruction x Problem Type	.395	1	.395	9.596	.003	.110
Instruction x Problem Type x AH4	.004	1	.004	.094	.760	.001
Error (Instruction * Problem Type)	3.209	78	.041			
Instruction x Complexity	.001	1	.001	.092	.762	.001
Instruction x Complexity x AH4	.001	1	.001	.096	.757	.001
Error (Instruction * Complexity)	.754	78	.010			
Problem x Complexity	.051	1	.051	5.709	.019	.068
Problem x Complexity x AH4	.009	1	.009	1.004	.319	.013
Error (Problem * Complexity)	.697	78	.009			
Instruction x Problem x Complexity	.014	1	.014	1.439	.234	.018
Instruction x Problem x Complexity x AH4	2.458E-005	1	2.458E-005	.003	.960	.000
Error (Instruction x Problem x Complexity)	.753	78	.010			

Note: AH4 group (high and low) as a between subjects factor

Table C4.14 Between subject effects on full accuracy data set – N = 80 – AH4

	SS	df	MS	F	*p*	η_p^2
Intercept	422.202	1	422.202	2810.566	.000	.973
AH4 group	1.760	1	1.760	11.718	.001	.131
Error	11.717	78	.150			

Table C4.15 Repeated measures ANOVA on excluded accuracy data set – N = 72 – CRT

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.919	1	2.919	25.783	.000	.269
Instruction x CRT (high or low group)	.455	1	.455	4.024	.049	.054
Error (Instruction)	7.924	70	.113			
Problem Type (Conflict/No-conflict)	5.794	1	5.794	57.050	.000	.449
Problem Type x CRT	.329	1	.329	3.235	.076	.044
Error (Problem Type)	7.109	70	.102			
Complexity (MP/Disjunctives)	2.648	1	2.648	39.322	.000	.360
Complexity x CRT	.040	1	.040	.591	.445	.008
Error (Complexity)	4.714	70	.067			
Instruction x Problem Type	2.029	1	2.029	21.610	.000	.236
Instruction x Problem Type x CRT	.215	1	.215	2.286	.135	.032
Error (Instruction * Problem Type)	6.572	70	.094			
Instruction x Complexity	.074	1	.074	1.914	.171	.027
Instruction x Complexity x CRT	.014	1	.014	.364	.549	.005
Error (Instruction * Complexity)	2.708	70	.039			
Problem x Complexity	.297	1	.297	7.526	.008	.097
Problem x Complexity x CRT	.009	1	.009	.240	.626	.003
Error (Problem * Complexity)	2.760	70	.039			
Instruction x Problem x Complexity	.098	1	.098	2.826	.097	.039
Instruction x Problem x Complexity x CRT	.010	1	.010	.300	.586	.004
Error (Instruction x Problem x Complexity)	2.423	70	.035			

Note: CRT group (high and low) as a between subjects factor

Table C4.16 Between subject effects on excluded accuracy data set – N = 72 – CRT

	SS	df	MS	F	*p*	η_p^2
Intercept	707.004	1	707.004	1161.962	.000	.943
CRT group	.249	1	.249	.409	.524	.006
Error	42.592	70	.608			

Table C4.17 Repeated measures ANOVA on full accuracy data set – N = 80 – CRT

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	.661	1	.661	15.506	.000	.166
Instruction x CRT (high or low group)	.224	1	.224	5.254	.025	.063
Error (Instruction)	3.327	78	.043			
Problem Type (Conflict/No-conflict)	2.732	1	2.732	54.824	.000	.413
Problem Type x CRT	.014	1	.014	.278	.600	.004
Error (Problem Type)	3.886	78	.050			
Complexity (MP/Disjunctives)	.356	1	.356	37.529	.000	.325
Complexity x CRT	.017	1	.017	1.748	.190	.022
Error (Complexity)	.739	78	.009			
Instruction x Problem Type	.504	1	.504	13.030	.001	.143
Instruction x Problem Type x CRT	.194	1	.194	5.024	.028	.061
Error (Instruction * Problem Type)	3.018	78	.039			
Instruction x Complexity	.001	1	.001	.055	.815	.001
Instruction x Complexity x CRT	.001	1	.001	.093	.761	.001
Error (Instruction * Complexity)	.754	78	.010			
Problem x Complexity	.051	1	.051	5.643	.020	.067
Problem x Complexity x CRT	.000	1	.000	.027	.870	.000
Error (Problem * Complexity)	.706	78	.009			
Instruction x Problem x Complexity	.011	1	.011	1.182	.280	.015
Instruction x Problem x Complexity x CRT	.001	1	.001	.139	.710	.002
Error (Instruction x Problem x Complexity)	.751	78	.010			

Note: CRT group (high and low) as a between subjects factor

Table C4.18 Between subject effects on full accuracy data set – N = 80 – CRT

	SS	df	MS	F	p	η_p^2
Intercept	404.363	1	404.363	2385.818	.000	.968
CRT group	.257	1	.257	1.519	.221	.019
Error	13.220	78	.169			

Table C4.19 Repeated measures ANOVA on excluded accuracy data set – N = 72 - AOT

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.514	1	2.514	21.325	.000	.234
Instruction x AOT (high or low group)	.127	1	.127	1.080	.302	.015
Error (Instruction)	8.252	70	.118			
Problem Type (Conflict/No-conflict)	6.643	1	6.643	63.053	.000	.474
Problem Type x AOT	.063	1	.063	.593	.444	.008
Error (Problem Type)	7.375	70	.105			
Complexity (MP/Disjunctives)	2.863	1	2.863	42.201	.000	.376
Complexity x AOT	.005	1	.005	.069	.794	.001
Error (Complexity)	4.749	70	.068			
Instruction x Problem Type	1.868	1	1.868	19.343	.000	.217
Instruction x Problem Type x AOT	.025	1	.025	.257	.614	.004
Error (Instruction * Problem Type)	6.761	70	.097			
Instruction x Complexity	.095	1	.095	2.445	.122	.034
Instruction x Complexity x AOT	.016	1	.016	.414	.522	.006
Error (Instruction * Complexity)	2.706	70	.039			
Problem x Complexity	.279	1	.279	7.094	.010	.092
Problem x Complexity x AOT	.016	1	.016	.396	.531	.006
Error (Problem * Complexity)	2.754	70	.039			
Instruction x Problem x Complexity	.086	1	.086	2.481	.120	.034
Instruction x Problem x Complexity x AOT	.008	1	.008	.236	.629	.003
Error (Instruction x Problem x Complexity)	2.425	70	.035			

Note: AOT (Active Open-minded Thinking) group (high and low) as a between subjects factor

Table C4.20 Between subject effects on excluded accuracy data set – N = 72 - AOT

	SS	df	MS	F	*p*	η_p^2
Intercept	724.296	1	724.296	1209.686	.000	.945
AOT group	.929	1	.929	1.551	.217	.022
Error	41.912	70	.599			

Table C4.21 Repeated measures ANOVA on full accuracy data set – N = 80 - AOT

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.527	1	.527	11.583	.001	.129
Instruction x AOT (high or low group)	.001	1	.001	.030	.863	.000
Error (Instruction)	3.550	78	.046			
Problem Type (Conflict/No-conflict)	2.941	1	2.941	60.430	.000	.437
Problem Type x AOT	.105	1	.105	2.153	.146	.027
Error (Problem Type)	3.796	78	.049			
Complexity (MP/Disjunctives)	.412	1	.412	42.574	.000	.353
Complexity x AOT	.000	1	.000	.043	.836	.001
Error (Complexity)	.755	78	.010			
Instruction x Problem Type	.388	1	.388	9.489	.003	.108
Instruction x Problem Type x AOT	.020	1	.020	.500	.482	.006
Error (Instruction * Problem Type)	3.192	78	.041			
Instruction x Complexity	.001	1	.001	.091	.764	.001
Instruction x Complexity x AOT	.002	1	.002	.189	.665	.002
Error (Instruction * Complexity)	.753	78	.010			
Problem x Complexity	.052	1	.052	5.757	.019	.069
Problem x Complexity x AOT	3.079E-006	1	3.079E-006	.000	.985	.000
Error (Problem * Complexity)	.706	78	.009			
Instruction x Problem x Complexity	.014	1	.014	1.419	.237	.018
Instruction x Problem x Complexity x AOT	.002	1	.002	.184	.669	.002
Error (Instruction x Problem x Complexity)	.751	78	.010			

Note: AOT (Active Open-minded Thinking) group (high and low) as a between subjects factor

Table C4.22 Between subject effects on full accuracy data set – N = 80 - AOT

	SS	df	MS	F	*p*	η_p^2
Intercept	421.895	1	421.895	2649.376	.000	.971
AOT group	1.056	1	1.056	6.634	.012	.078
Error	12.421	78	.159			

Experiment 5

Randomness

Table C5.1 MANOVA on the Randomness Indices across the 3 Primary task conditions –N = 108

	Value	F	df	Error df	*p*	η_p^2
Intercept	.009	3927.615	3	103.00	.000	.991
Primary Task (Belief/Logic and baseline)	.747	5.386	6	206.000	.000	.136

Table C5.2 ANOVA measuring the effect of the Primary task on 3 Randomness Indices N = 108

		SS	df	MS	F	*p*	η_p^2
Intercept	Redundancy	.080	1	.080	148.275	.000	.585
	Adjacency	11.745	1	11.745	1307.550	.000	.926
	RNG	35.593	1	35.593	9177.353	.000	.989
Primary Task	Redundancy	.019	2	.009	17.316	.000	.248
	Adjacency	.008	2	.004	.458	.634	.009
	RNG	.000	2	9.259E-005	.024	.976	.000
Error	Redundancy	.057	105	.001			
	Adjacency	.943	105	.009			
	RNG	.407	105	.004			
Total	Redundancy	.155	108				
	Adjacency	12.697	108				
	RNG	36.000	108				

Table C5.3 Post Hoc Analysis (LSD) of Primary task on Randomness indices – N = 108

Dependent Variable	(I)Primary Task Instruction	(J)Primary Task Instruction	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Upper Bound	Lower Bound
Redundancy	Baseline	Belief	-.026646*	.0054694	.000	-.037491	-.015801
		Logic	-.028960*	.0054694	.000	-.039805	-.018115
	Belief	Baseline	.026646*	.0054694	.000	.015801	.037491
		Logic	-.002314	.0054694	.673	-.013159	.008531
	Logic	Baseline	.028960*	.0054694	.000	.018115	.039805
		Belief	.002314	.0054694	.673	-.008531	.013159
Adjacency	Baseline	Belief	-.018422	.0223392	.411	-.062717	.025872
		Logic	-.018594	.0223392	.407	-.062889	.025700
	Belief	Baseline	.018422	.0223392	.411	-.025872	.062717
		Logic	-.000172	.0223392	.994	-.044467	.044122
	Logic	Baseline	.018594	.0223392	.407	-.025700	.062889
		Belief	.000172	.0223392	.994	-.044122	.044467
RNG	Baseline	Belief	.002778	.0146786	.850	-.026327	.031883
		Logic	.000000	.0146786	1.000	-.029105	.029105
	Belief	Baseline	-.002778	.0146786	.850	-.031883	.026327
		Logic	-.002778	.0146786	.850	-.031883	.026327
	Logic	Baseline	.000000	.0146786	1.000	-.029105	.029105
		Belief	.002778	.0146786	.850	-.026327	.031883

Table C5.4 RG Calc Output Table for RNG Group

Participant Number	Sample Size (minus outliers)	Seconds per Digit (with outliers)	R	A (ascending pairs)	A (decending pairs)	A (combined)	RNG	NSQ	TPI	Runs	RG Mean	RG Median	RG Mode	Most frequent digit	Least Frequent digit	Ordering (Belief 1st B1) or Logic 1st L1)
3	677	1.3	2.95%	18.32%	15.36%	33.68%	0.6	10.00%	94.00%	1.2	9	7	4	1	6	B1
4	1000	1	1.20%	18.50%	14.80%	33.30%	0.6	3.80%	90.00%	1.1	9	7	7	5	6	L1
5	758	1.2	2.60%	23.48%	15.70%	39.18%	0.6	6.00%	89.00%	1.3	9	7	5	2	7	B1
6	368	1.7	0.80%	17.66%	25.82%	43.48%	0.5	6.30%	73.00%	1.5	9	7	6	3	6	L1
9	846	1.2	1.80%	14.89%	14.07%	28.96%	0.6	5.00%	91.00%	0.6	9	7	5	4	8	B1
11	577	1.2	3.80%	17.04%	18.43%	35.48%	0.6	10.00%	90.00%	1.2	9	6	4	8	1	L1
13	699	0.76	5.40%	37.44%	8.14%	45.58%	0.6	25.00%	75.00%	2.3	9	6	6	1	9	B1
16	826	1	4.80%	25.42%	19.13%	44.55%	0.6	8.00%	88.00%	1.3	9	7	2	2	5	L1
18	534	1.5	1.50%	25.66%	13.30%	38.95%	0.5	8.80%	74.00%	1.8	9	7	5	7	1	L1
19	335	1.7	1.40%	9.58%	11.38%	20.96%	0.5	15.00%	96.00%	0.8	9	7	7	4	2	L1
20	699	1.6	6.50%	13.30%	15.74%	29.04%	0.6	10.00%	83.00%	1.2	9	6	3	2	3	B1
21	893	1	2.70%	16.57%	11.31%	27.88%	0.6	7.50%	87.00%	1.3	9	7	5	7	2	B1
22	779	0.97	1.90%	25.35%	15.19%	40.54%	0.6	1.25%	83.00%	1.8	9	7	4	2	9	B1
23	932	1	10.00%	10.62%	9.33%	19.96%	0.6	6.30%	79.00%	1.2	9	5	1	1	8	L1
29	734	1	3.00%	27.05%	10.25%	37.30%	0.6	3.60%	92.00%	1.6	9	7	6	4	6	B1
32	1000	1	2.20%	18.90%	17.00%	35.90%	0.6	6.30%	87.00%	1.3	9	7	5	2	3	L1
33	550	1.3	0.70%	16.73%	16.18%	32.91%	0.5	8.80%	85.00%	1.3	9	7	5	6	5	B1
36	383	1.2	1.20%	17.30%	17.78%	35.08%	0.6	3.00%	84.00%	1.2	9	7	6	2	6	L1
38	407	0.8	0.20%	7.37%	4.91%	12.29%	0.7	35.00%	66.00%	3	9	9	9	8	4,5,7	B1
40	818	1	0.60%	23.59%	20.42%	44.01%	0.6	3.80%	84.00%	1.5	9	7	4	3	9	L1
42	561	1	2.60%	37.19%	16.73%	53.91%	0.6	20.00%	70.00%	3	9	7	9	2	1	B1
44	636	1	3.00%	21.70%	9.28%	30.97%	0.6	6.30%	95.00%	1	9	7	6	3	6	L1
46	1000	0.95	0.60%	13.10%	13.00%	26.10%	0.6	1.30%	93.00%	0.9	9	8	6	7	8	B1
48	582	1	4.20%	18.35%	15.61%	33.96%	0.6	8.80%	70.00%	1.1	9	7	1	8	4	L1
50	365	1.4	4.00%	49.44%	8.71%	58.15%	0.61	20.00%	70.00%	4	9	7	5	4	1	L1
51	706	0.96	0.40%	9.35%	11.33%	20.68%	0.5	8.80%	103.00%	0.8	9	8	6	1	5	B1
54	820	1	8.30%	12.58%	10.50%	23.08%	0.6	14.00%	98.00%	0.99	9	6	4	5	4	B1
55	740	1	3.20%	19.73%	16.49%	36.22%	0.6	10.00%	93.00%	1.2	9	7	6	3	2	L1
57	1000	1	1.50%	14.50%	12.80%	27.30%	0.6	5.00%	93.00%	1.2	9	8	6	7	2	L1
58	806	1.4	7.60%	13.79%	8.94%	22.73%	0.6	10.00%	82.00%	1.1	9	6	1	7	6	B1
59	497	1.6	2.80%	20.72%	17.91%	38.63%	0.5	11.00%	86.00%	1.4	9	7	5	1	5	B1
60	305	1.8	7.30%	15.41%	9.51%	24.92%	0.5	25.00%	88.00%	1.1	9	5	4	5	3	L1
61	568	1	2.70%	19.72%	14.08%	33.80%	0.5	3.80%	79.00%	1.6	9	7	5	2	6	B1
62	486	1	5.00%	21.19%	15.23%	36.42%	0.6	11.00%	60.00%	2	9	6	1	2	7	L1
63	628	0.99	2.00%	17.20%	25.32%	42.52%	0.6	7.50%	83.00%	1.3	9	7	6	3	7	B1
64	764	1.2	7.50%	22.91%	13.22%	36.13%	0.6	2.50%	77.00%	1.5	9	6	1	1	8	L1

Accuracy

Table C5.5 Repeated measures ANOVA on excluded data set – N = 61

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.659	1	1.659	14.359	.000	.196
Instruction x Condition (RNG or Control)	.843	1	.843	7.295	.009	.110
Error (Instruction)	6.816	59	.116			
Problem Type (Conflict/No-conflict)	4.286	1	4.286	40.138	.000	.405
Problem Type x Condition	.054	1	.054	.505	.480	.008
Error (Problem Type)	6.301	59	.107			
Complexity (MP/Disjunctives)	1.526	1	1.526	22.195	.000	.273
Complexity x Condition	.130	1	.130	1.888	.175	.031
Error (Complexity)	4.056	59	.069			
Instruction x Problem Type	.619	1	.619	9.778	.003	.142
Instruction x Problem Type x Condition	.029	1	.029	.454	.503	.008
Error (Instruction * Problem Type)	3.737	59	.063			
Instruction x Complexity	.062	1	.062	1.071	.305	.018
Instruction x Complexity x Condition	.074	1	.074	1.286	.261	.021
Error (Instruction * Complexity)	3.400	59	.058			
Problem x Complexity	.036	1	.036	1.374	.246	.023
Problem x Complexity x Condition	.050	1	.050	1.932	.170	.032
Error (Problem * Complexity)	1.536	59	.026			
Instruction x Problem x Complexity	.028	1	.028	.512	.477	.009
Instruction x Problem x Complexity x Condition	.006	1	.006	.103	.750	.002
Error (Instruction x Problem x Complexity)	3.231	59	.055			

Note: Condition is a between subjects factor

Table C5.6 Between subject effects on excluded data set – N = 61

	SS	df	MS	F	*p*	η_p^2
Intercept	638.864	1	638.864	1262.663	.000	.955
Condition	4.313	1	4.313	8.525	.005	.126
Error	29.852	59	.506			

Table C5.7 Repeated measures ANOVA on full data set – N = 71

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	.478	1	.478	10.253	.002	.129
Instruction x Condition (RNG or Control)	.011	1	.011	.245	.622	.004
Error (Instruction)	3.215	69	.047			
Problem Type (Conflict/No-conflict)	3.113	1	3.113	53.074	.000	.435
Problem Type x Condition	.001	1	.001	.016	.901	.000
Error (Problem Type)	4.047	69	.059			
Complexity (MP/Disjunctives)	.213	1	.213	17.154	.000	.199
Complexity x Condition	.038	1	.038	3.077	.084	.043
Error (Complexity)	.856	69	.012			
Instruction x Problem Type	.452	1	.452	10.353	.002	.130
Instruction x Problem Type x Condition	.012	1	.012	.270	.605	.004
Error (Instruction * Problem Type)	3.013	69	.044			
Instruction x Complexity	.005	1	.005	.367	.547	.005
Instruction x Complexity x Condition	.021	1	.021	1.593	.211	.023
Error (Instruction * Complexity)	.889	69	.013			
Problem x Complexity	.005	1	.005	.805	.373	.012
Problem x Complexity x Condition	.022	1	.022	3.243	.076	.045
Error (Problem * Complexity)	.469	69	.007			
Instruction x Problem x Complexity	.004	1	.004	.328	.568	.005
Instruction x Problem x Complexity x Condition	.006	1	.006	.515	.475	.007
Error (Instruction x Problem x Complexity)	.792	69	.011			

Note: Condition *is a between subjects factor*

Table C5.8 Between subject effects on full data set – N = 71

	SS	df	MS	F	p	η_p^2
Intercept	376.386	1	376.386	2435.064	.000	.972
Condition	1.118	1	1.118	7.232	.009	.095
Error	10.665	69	.155			

<u>Follow Up Analysis</u>

Table C5.9 Examining the effect of condition on Belief instruction across early blocks (1 &2) – excluded data set – N = 61

	SS	df	MS	F	p	η_p^2
Intercept	310.932	1	310.932	913.999	.000	.939
Condition	1.001	1	1.001	2.943	.092	.048
Error	20.071	59	.340			

Table C5.10 Examining the effect of condition on Belief instruction across later blocks (3 &4) – excluded data set – N = 61

	SS	df	MS	F	*p*	η_p^2
Intercept	306.871	1	306.871	706.326	.000	.923
Condition	1.027	1	1.027	2.363	.130	.039
Error	25.633	59	.434			

Latencies

Table C5.11 Repeated measures ANOVA on excluded data set – N = 61

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	36750924.696	1	36750924.696	5.786	.019	.089
Instruction x Condition (RNG or Control)	16284021.057	1	16284021.057	2.564	.115	.042
Error (Instruction)	374763452.919	59	6351922.931			
Problem Type (Conflict/No-conflict)	196103.463	1	196103.463	.120	.731	.002
Problem Type x Condition	8470462.479	1	8470462.479	5.170	.027	.081
Error (Problem Type)	96668264.201	59	1638445.156			
Complexity (MP/Disjunctives)	289359950.064	1	289359950.064	108.661	.000	.648
Complexity x Condition	2694863.670	1	2694863.670	1.012	.319	.017
Error (Complexity)	157115072.551	59	2662967.331			
Instruction x Problem Type	7796600.776	1	7796600.776	2.902	.094	.047
Instruction x Problem Type x Condition	3229365.662	1	3229365.662	1.202	.277	.020
Error (Instruction * Problem Type)	158491033.560	59	2686288.704			
Instruction x Complexity	443697.311	1	443697.311	.284	.596	.005
Instruction x Complexity x Condition	2880771.409	1	2880771.409	1.847	.179	.030
Error (Instruction * Complexity)	92030819.468	59	1559844.398			
Problem x Complexity	93862.337	1	93862.337	.054	.817	.001
Problem x Complexity x Condition	1752.993	1	1752.993	.001	.975	.000
Error (Problem * Complexity)	101989945.785	59	1728643.149			
Instruction x Problem x Complexity	39239.311	1	39239.311	.024	.877	.000
Instruction x Problem x Complexity x Condition	853038.065	1	853038.065	.521	.473	.009
Error (Instruction x Problem x Complexity)	96567979.911	59	1636745.422			

Note: Condition is a between subjects factor

Table C5.12 Between subject effects on excluded data set – N = 61

	SS	df	MS	F	p	η_p^2
Intercept	16030578053.285	1	16030578053.285	878.593	.000	.937
Condition	302806557.219	1	302806557.219	16.596	.000	.220
Error	1076498531.674	59	18245737.825			

Table C5.13 Repeated measures ANOVA on full data set – N = 71

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	21671446.651	1	21671446.651	1.689	.198	.024
Instruction x Condition (RNG or Control)	13894197.292	1	13894197.292	1.083	.302	.015
Error (Instruction)	885185347.374	69	12828773.150			
Problem Type (Conflict/No-conflict)	23903.605	1	23903.605	.015	.902	.000
Problem Type x Condition	4976603.992	1	4976603.992	3.157	.080	.044
Error (Problem Type)	108777815.455	69	1576490.079			
Complexity (MP/Disjunctives)	365786420.074	1	365786420.074	77.553	.000	.529
Complexity x Condition	14960521.448	1	14960521.448	3.172	.079	.044
Error (Complexity)	325443713.964	69	4716575.565			
Instruction x Problem Type	4647749.209	1	4647749.209	2.274	.136	.032
Instruction x Problem Type x Condition	4816217.766	1	4816217.766	2.357	.129	.033
Error (Instruction * Problem Type)	141013679.372	69	2043676.513			
Instruction x Complexity	193286.358	1	193286.358	.078	.781	.001
Instruction x Complexity x Condition	35271.591	1	35271.591	.014	.905	.000
Error (Instruction * Complexity)	170450267.624	69	2470293.734			
Problem x Complexity	116152.255	1	116152.255	.034	.854	.000
Problem x Complexity x Condition	2916122.755	1	2916122.755	.856	.358	.012
Error (Problem * Complexity)	235113190.481	69	3407437.543			
Instruction x Problem x Complexity	1112744.294	1	1112744.294	.300	.586	.004
Instruction x Problem x Complexity x Condition	4900111.301	1	4900111.301	1.321	.254	.019
Error (Instruction x Problem x Complexity)	255921990.344	69	3709014.353			

Note: Condition is a between subjects factor

Table C5.14 Between subject effects on full data set – N = 71

	SS	df	MS	F	*p*	η_p^2
Intercept	20724930175.794	1	20724930175.794	617.711	.000	.900
Condition	785664425.618	1	785664425.618	23.417	.000	.253
Error	2315032267.372	69	33551192.281			

Follow Up Analysis

Table C5.15 Repeated measures ANOVA on Control Condition only - excluded accuracy data set – N = 30

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2142162.907	1	2142162.907	.760	.390	.025
Instruction x Condition	.000	0	.	.	.	.000
Error (Instruction)	84542285.968	30	2818076.199			
Problem Type (Conflict/No-conflict)	5804514.036	1	5804514.036	5.522	.026	.155
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	31536476.339	30	1051215.878			
Complexity (MP/Disjunctives)	119666521.940	1	119666521.940	74.805	.000	.714
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	47991630.935	30	1599721.031			
Instruction x Problem Type	530025.101	1	530025.101	.420	.522	.014
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	37824110.774	30	1260803.692			
Instruction x Complexity	513695.036	1	513695.036	.528	.473	.017
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	29177705.339	30	972590.178			
Problem x Complexity	71162.907	1	71162.907	.091	.766	.003
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	23589211.968	30	786307.066			
Instruction x Problem x Complexity	287028.101	1	287028.101	.213	.648	.007
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	40457423.274	30	1348580.776			

Table C5.16 Repeated measures ANOVA on RNG Condition only - excluded accuracy data set – N = 29

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	43070801.004	1	43070801.004	4.284	.047	.129
Instruction x Condition	.000	0	.	.	.	.000
Error (Instruction)	291570026.371	29	10054138.840			
Problem Type (Conflict/No-conflict)	5551954.204	1	5551954.204	2.514	.124	.080
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	64049592.171	29	2208606.627			
Complexity (MP/Disjunctives)	156983227.538	1	156983227.538	54.587	.000	.653
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	83398506.838	29	2875810.581			
Instruction x Problem Type	14461423.204	1	14461423.204	2.990	.094	.093
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	140251675.671	29	4836264.678			
Instruction x Complexity	1796605.104	1	1796605.104	.982	.330	.033
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	53077748.771	29	1830267.199			
Problem x Complexity	109953.204	1	109953.204	.038	.847	.001
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	84415753.671	29	2910888.058			
Instruction x Problem x Complexity	268068.504	1	268068.504	.113	.739	.004
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	68554218.871	29	2363938.582			

Experiment 7

Accuracy

Table C6.1 Repeated measures ANOVA on N-Back Instruction only - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.090	1	.090	2.243	.139	.030
Problem Type x Condition	.066	1	.066	1.654	.202	.022
Error (Problem Type)	2.932	73	.040			
Complexity (MP/Disjunctives)	.237	1	.237	4.094	.047	.053
Complexity x Condition	.237	1	.237	4.105	.046	.053
Error (Complexity)	4.222	73	.058			
Problem x Complexity	.006	1	.006	.155	.695	.002
Problem x Complexity x Condition	.198	1	.198	5.257	.025	.067
Error (Problem * Complexity)	2.744	73	.038			

Note: Condition is a between subjects factor

Table C6.2 Between subject effects on N-Back Instruction only - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Intercept	308.537	1	308.537	905.456	.000	.925
Condition	.650	1	.650	1.908	.171	.025
Error	24.875	73	.341			

Table C6.3 Repeated measures ANOVA on N-Back Instruction only - full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.046	1	.046	3.257	.075	.040
Problem Type x Condition	.046	1	.046	3.257	.075	.040
Error (Problem Type)	1.115	79	.014			
Complexity (MP/Disjunctives)	.107	1	.107	5.624	.020	.066
Complexity x Condition	.098	1	.098	5.161	.026	.061
Error (Complexity)	1.500	79	.019			
Problem x Complexity	.025	1	.025	2.217	.141	.027
Problem x Complexity x Condition	.065	1	.065	5.692	.019	.067
Error (Problem * Complexity)	.905	79	.011			

Note: Condition is a between subjects factor

Table C6.4 Between subject effects on N-Back Instruction only - full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Intercept	201.102	1	201.102	2052.177	.000	.963
Condition	.166	1	.166	1.696	.197	.021
Error	7.742	79	.098			

Follow Up Analysis

Table C6.5 Repeated measures ANOVA on Experimental Condition only - excluded accuracy data set – N = 37

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.154	1	.154	3.300	.078	.084
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	1.676	36	.047			
Complexity (MP/Disjunctives)	.468	1	.468	7.933	.008	.181
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	2.124	36	.059			
Problem x Complexity	.134	1	.134	2.434	.127	.063
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.979	36	.055			

Table C6.6 Repeated measures ANOVA on Control Condition only - excluded accuracy data set – N = 38

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.001	1	.001	.027	.871	.001
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	1.256	37	.034			
Complexity (MP/Disjunctives)	2.141E-007	1	2.141E-007	.000	.998	.000
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	2.099	37	.057			
Problem x Complexity	.069	1	.069	3.326	.076	.082
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	.764	37	.021			

Table C6.7 Repeated measures ANOVA on Belief & Logic Instruction - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.935	1	2.935	35.460	.000	.327
Instruction x Condition (Experimental or Control)	.337	1	.337	4.073	.047	.053
Error (Instruction)	6.042	73	.083			
Problem Type (Conflict/No-conflict)	2.697	1	2.697	33.004	.000	.311
Problem Type x Condition	.110	1	.110	1.350	.249	.018
Error (Problem Type)	5.966	73	.082			
Complexity (MP/Disjunctives)	2.782	1	2.782	37.279	.000	.338
Complexity x Condition	.002	1	.002	.028	.868	.000
Error (Complexity)	5.447	73	.075			
Instruction x Problem Type	.823	1	.823	10.899	.001	.130
Instruction x Problem Type x Condition	.251	1	.251	3.324	.072	.044
Error (Instruction * Problem Type)	5.514	73	.076			
Instruction x Complexity	.286	1	.286	4.586	.036	.059
Instruction x Complexity x Condition	.000	1	.000	.003	.957	.000
Error (Instruction * Complexity)	4.553	73	.062			
Problem x Complexity	.412	1	.412	14.367	.000	.164
Problem x Complexity x Condition	.067	1	.067	2.346	.130	.031
Error (Problem * Complexity)	2.092	73	.029			
Instruction x Problem x Complexity	.134	1	.134	3.656	.060	.048
Instruction x Problem x Complexity x Condition	.072	1	.072	1.954	.166	.026
Error (Instruction x Problem x Complexity)	2.673	73	.037			

Note: Condition is a between subjects factor

Table C6.8 Between subject effects on Belief & Logic Instruction - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Intercept	1008.513	1	1008.513	2841.077	.000	.975
Condition	.242	1	.242	.683	.411	.009
Error	25.913	73	.355			

Table C6.9 Repeated measures ANOVA on Belief & Logic Instruction - full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.681	1	.681	20.499	.000	.206
Instruction x Condition (Experimental or Control)	.135	1	.135	4.066	.047	.049
Error (Instruction)	2.626	79	.033			
Problem Type (Conflict/No-conflict)	1.337	1	1.337	26.680	.000	.252
Problem Type x Condition	.045	1	.045	.892	.348	.011
Error (Problem Type)	3.960	79	.050			
Complexity (MP/Disjunctives)	.441	1	.441	39.061	.000	.331
Complexity x Condition	.007	1	.007	.590	.445	.007
Error (Complexity)	.891	79	.011			
Instruction x Problem Type	.501	1	.501	16.114	.000	.169
Instruction x Problem Type x Condition	.116	1	.116	3.722	.057	.045
Error (Instruction * Problem Type)	2.454	79	.031			
Instruction x Complexity	.081	1	.081	6.510	.013	.076
Instruction x Complexity x Condition	.002	1	.002	.145	.705	.002
Error (Instruction * Complexity)	.987	79	.012			
Problem x Complexity	.136	1	.136	16.001	.000	.168
Problem x Complexity x Condition	.008	1	.008	.908	.344	.011
Error (Problem * Complexity)	.673	79	.009			
Instruction x Problem x Complexity	.047	1	.047	6.633	.012	.077
Instruction x Problem x Complexity x Condition	7.821E-006	1	7.821E-006	.001	.974	.000
Error (Instruction x Problem x Complexity)	.557	79	.007			

Note: Condition is a between subjects factor

Table C6.10 Between subject effects on Belief & Logic Instruction - full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Intercept	512.743	1	512.743	4398.477	.000	.982
Condition	.018	1	.018	.151	.699	.002
Error	9.209	79	.117			

Latencies

Table C6.11 Repeated measures ANOVA on N-Back Instruction only - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	2673853.144	1	2673853.144	4.746	.033	.061
Problem Type x Condition	1256336.770	1	1256336.770	2.230	.140	.030
Error (Problem Type)	41130685.376	73	563434.046			
Complexity (MP/Disjunctives)	867923.999	1	867923.999	1.594	.211	.021
Complexity x Condition	204161.333	1	204161.333	.375	.542	.005
Error (Complexity)	39751775.787	73	544544.874			
Problem x Complexity	400854.684	1	400854.684	.574	.451	.008
Problem x Complexity x Condition	835594.044	1	835594.044	1.196	.278	.016
Error (Problem * Complexity)	51003938.503	73	698684.089			

Note: Condition is a between subjects factor

Table C6.12 Between subject effects on N-Back Instruction only - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Intercept	3154081627.589	1	3154081627.589	595.575	.000	.891
Condition	342605.456	1	342605.456	.065	.800	.001
Error	386597984.331	73	5295862.799			

Table C6.13 Repeated measures ANOVA on N-Back Instruction only - full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	1550575.587	1	1550575.587	1.231	.271	.015
Problem Type x Condition	1550028.994	1	1550028.994	1.230	.271	.015
Error (Problem Type)	99548391.419	79	1260106.220			
Complexity (MP/Disjunctives)	2079779.895	1	2079779.895	2.045	.157	.025
Complexity x Condition	339934.414	1	339934.414	.334	.565	.004
Error (Complexity)	80336753.889	79	1016920.935			
Problem x Complexity	1961908.684	1	1961908.684	2.193	.143	.027
Problem x Complexity x Condition	1861165.129	1	1861165.129	2.081	.153	.026
Error (Problem * Complexity)	70668198.396	79	894534.157			

Note: Condition is a between subjects factor

Table C6.14 Between subject effects on N-Back Instruction only - full data set – N = 81

	SS	df	MS	F	p	η_p^2
Intercept	4147417057.135	1	4147417057.135	430.642	.000	.845
Condition	11007188.122	1	11007188.122	1.143	.288	.014
Error	760831637.588	79	9630780.223			

Table C6.15 Repeated measures ANOVA on Belief & Logic Instruction - excluded data set – N = 75

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	11536018.985	1	11536018.985	5.628	.020	.072
Instruction x Condition (Experimental or Control)	2023424.065	1	2023424.065	.987	.324	.013
Error (Instruction)	149628853.359	73	2049710.320			
Problem Type (Conflict/No-conflict)	21812799.061	1	21812799.061	14.403	.000	.165
Problem Type x Condition	81827.501	1	81827.501	.054	.817	.001
Error (Problem Type)	110558631.082	73	1514501.796			
Complexity (MP/Disjunctives)	1631774.347	1	1631774.347	1.352	.249	.018
Complexity x Condition	2420921.827	1	2420921.827	2.006	.161	.027
Error (Complexity)	88116938.796	73	1207081.353			
Instruction x Problem Type	1200619.930	1	1200619.930	.927	.339	.013
Instruction x Problem Type x Condition	80840.103	1	80840.103	.062	.803	.001
Error (Instruction * Problem Type)	94565845.480	73	1295422.541			
Instruction x Complexity	1503017.796	1	1503017.796	1.043	.311	.014
Instruction x Complexity x Condition	4134750.396	1	4134750.396	2.868	.095	.038
Error (Instruction * Complexity)	105240384.547	73	1441649.103			
Problem x Complexity	1028600.815	1	1028600.815	.990	.323	.013
Problem x Complexity x Condition	1814046.721	1	1814046.721	1.746	.191	.023
Error (Problem * Complexity)	75858869.529	73	1039162.596			
Instruction x Problem x Complexity	2198319.065	1	2198319.065	1.574	.214	.021
Instruction x Problem x Complexity x Condition	923287.905	1	923287.905	.661	.419	.009
Error (Instruction x Problem x Complexity)	101923165.678	73	1396207.749			

Note: Condition is a between subjects factor

Table C6.16 Between subject effects on Belief & Logic Instruction - excluded data set – N = 75

	SS	df	MS	F	*p*	η_p^2
Intercept	27723381354.280	1	27723381354.280	1381.670	.000	.950
Condition	3075157.760	1	3075157.760	.153	.697	.002
Error	1464753738.263	73	20065119.702			

Table C6.17 Repeated measures ANOVA on Belief & Logic Instruction - full data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2762883.311	1	2762883.311	.692	.408	.009
Instruction x Condition (Experimental or Control)	3520792.941	1	3520792.941	.882	.351	.011
Error (Instruction)	315341543.247	79	3991665.104			
Problem Type (Conflict/No-conflict)	22907127.562	1	22907127.562	9.718	.003	.110
Problem Type x Condition	3329663.117	1	3329663.117	1.413	.238	.018
Error (Problem Type)	186217407.633	79	2357182.375			
Complexity (MP/Disjunctives)	10948157.080	1	10948157.080	3.803	.055	.046
Complexity x Condition	134546.562	1	134546.562	.047	.829	.001
Error (Complexity)	227439129.516	79	2878976.323			
Instruction x Problem Type	867421.874	1	867421.874	.312	.578	.004
Instruction x Problem Type x Condition	3569141.207	1	3569141.207	1.282	.261	.016
Error (Instruction * Problem Type)	219982934.018	79	2784594.101			
Instruction x Complexity	941542.541	1	941542.541	.236	.629	.003
Instruction x Complexity x Condition	8099741.676	1	8099741.676	2.028	.158	.025
Error (Instruction * Complexity)	315548941.518	79	3994290.399			
Problem x Complexity	112032.805	1	112032.805	.060	.807	.001
Problem x Complexity x Condition	503195.398	1	503195.398	.269	.605	.003
Error (Problem * Complexity)	147683241.828	79	1869408.124			
Instruction x Problem x Complexity	803836.241	1	803836.241	.227	.635	.003
Instruction x Problem x Complexity x Condition	7792211.106	1	7792211.106	2.204	.142	.027
Error (Instruction x Problem x Complexity)	279280021.286	79	3535190.143			

Note: Condition is a between subjects factor

Table C6.18 Between subject effects on Belief & Logic Instruction - full data set – N = 81

	SS	df	MS	F	p	η_p^2
Intercept	32333320636.290	1	32333320636.290	1036.812	.000	.929
Condition	1435482.981	1	1435482.981	.046	.831	.001
Error	2463641002.213	79	31185329.142			

Individual Differences

Table C6.19 Repeated measures ANOVA on excluded accuracy data set – N = 75

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	2.404	1	2.404	27.869	.000	.276
Instruction x CRT (high or low group)	.082	1	.082	.946	.334	.013
Error (Instruction)	6.298	73	.086			
Problem Type (Conflict/No-conflict)	2.089	1	2.089	25.767	.000	.261
Problem Type x CRT	.159	1	.159	1.964	.165	.026
Error (Problem Type)	5.918	73	.081			
Complexity (MP/Disjunctives)	2.210	1	2.210	30.196	.000	.293
Complexity x CRT	.108	1	.108	1.470	.229	.020
Error (Complexity)	5.342	73	.073			
Instruction x Problem Type	.475	1	.475	6.402	.014	.081
Instruction x Problem Type x CRT	.348	1	.348	4.683	.034	.060
Error (Instruction * Problem Type)	5.417	73	.074			
Instruction x Complexity	.136	1	.136	2.282	.135	.030
Instruction x Complexity x CRT	.210	1	.210	3.526	.064	.046
Error (Instruction * Complexity)	4.344	73	.060			
Problem x Complexity	.253	1	.253	9.039	.004	.110
Problem x Complexity x CRT	.116	1	.116	4.151	.045	.054
Error (Problem * Complexity)	2.043	73	.028			
Instruction x Problem x Complexity	.128	1	.128	3.399	.069	.044
Instruction x Problem x Complexity x CRT	.000	1	.000	.009	.923	.000
Error (Instruction x Problem x Complexity)	2.745	73	.038			

Note: CRT group (high and low) as a between subjects factor

Table C6.20 Between subject effects on excluded accuracy data set – N = 75

	SS	df	MS	F	p	η_p^2
Intercept	943.319	1	943.319	2901.901	.000	.975
CRT group	2.426	1	2.426	7.462	.008	.093
Error	23.730	73	.325			

Table C6.21 Repeated measures ANOVA on full accuracy data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.506	1	.506	14.832	.000	.158
Instruction x CRT (high or low group)	.064	1	.064	1.866	.176	.023
Error (Instruction)	2.697	79	.034			
Problem Type (Conflict/No-conflict)	.966	1	.966	19.798	.000	.200
Problem Type x CRT	.151	1	.151	3.087	.083	.038
Error (Problem Type)	3.854	79	.049			
Complexity (MP/Disjunctives)	.375	1	.375	33.135	.000	.295
Complexity x CRT	.003	1	.003	.300	.585	.004
Error (Complexity)	.894	79	.011			
Instruction x Problem Type	.334	1	.334	10.695	.002	.119
Instruction x Problem Type x CRT	.103	1	.103	3.307	.073	.040
Error (Instruction * Problem Type)	2.466	79	.031			
Instruction x Complexity	.054	1	.054	4.386	.039	.053
Instruction x Complexity x CRT	.015	1	.015	1.251	.267	.016
Error (Instruction * Complexity)	.973	79	.012			
Problem x Complexity	.101	1	.101	11.960	.001	.131
Problem x Complexity x CRT	.010	1	.010	1.238	.269	.015
Error (Problem * Complexity)	.670	79	.008			
Instruction x Problem x Complexity	.042	1	.042	5.960	.017	.070
Instruction x Problem x Complexity x CRT	4.895E-006	1	4.895E-006	.001	.979	.000
Error (Instruction x Problem x Complexity)	.557	79	.007			

Note: CRT group (high and low) as a between subjects factor

Table C6.22 Between subject effects on full accuracy data set – N = 81

	SS	df	MS	F	*p*	η_p^2
Intercept	472.409	1	472.409	4224.626	.000	.982
CRT group	.393	1	.393	3.513	.065	.043
Error	8.834	79	.112			

__Follow Up Analysis__

Table C6.23 Repeated measures ANOVA on High CRT group only - excluded accuracy data set – N = 26

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.612	1	.612	14.384	.001	.365
Instruction x CRT (high or low group)	.000	0	.	.	.	.000
Error (Instruction)	1.064	25	.043			
Problem Type (Conflict/No-conflict)	.419	1	.419	10.887	.003	.303
Problem Type x CRT	.000	0	.	.	.	.000
Error (Problem Type)	.962	25	.038			
Complexity (MP/Disjunctives)	.514	1	.514	9.747	.004	.281
Complexity x CRT	.000	0	.	.	.	.000
Error (Complexity)	1.317	25	.053			
Instruction x Problem Type	.004	1	.004	.110	.743	.004
Instruction x Problem Type x CRT	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	.864	25	.035			
Instruction x Complexity	.003	1	.003	.072	.791	.003
Instruction x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction * Complexity)	1.067	25	.043			
Problem x Complexity	.010	1	.010	.467	.500	.018
Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Problem * Complexity)	.538	25	.022			
Instruction x Problem x Complexity	.054	1	.054	2.206	.150	.081
Instruction x Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	.614	25	.025			

Table C6.24 Repeated measures ANOVA on Low CRT group only - excluded accuracy data set – N = 49

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.431	1	2.431	22.300	.000	.317
Instruction x CRT (high or low group)	.000	0	.	.	.	.000
Error (Instruction)	5.234	48	.109			
Problem Type (Conflict/No-conflict)	2.453	1	2.453	23.759	.000	.331
Problem Type x CRT	.000	0	.	.	.	.000
Error (Problem Type)	4.956	48	.103			
Complexity (MP/Disjunctives)	2.374	1	2.374	28.317	.000	.371
Complexity x CRT	.000	0	.	.	.	.000
Error (Complexity)	4.024	48	.084			
Instruction x Problem Type	1.179	1	1.179	12.431	.001	.206
Instruction x Problem Type x CRT	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	4.554	48	.095			
Instruction x Complexity	.493	1	.493	7.218	.010	.131
Instruction x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction * Complexity)	3.276	48	.068			
Problem x Complexity	.513	1	.513	16.376	.000	.254
Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.505	48	.031			
Instruction x Problem x Complexity	.083	1	.083	1.863	.179	.037
Instruction x Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	2.130	48	.044			

Table C6.25 Correlations Matrix for Global Span (GS) scores against both Instruction and Problem Types – N = 75

		Global Span Score	Belief	Belief No-conflict	Belief Conflict	Logic	Logic No-conflict	Logic Conflict
GS Scores	P- Corr	1	.330**	.204	.315**	.177	.171	.167
	Sig.		.004	.079	.006	.128	.141	.152
	N	75	75	75	75	75	75	75
Belief	P- Corr	.330**	1	.563**	.971**	.457**	.488**	.390**
	Sig.	.004		.000	.000	.000	.000	.001
	N	75	75	75	75	75	75	75
Belief No-conflict	P- Corr	.204	.563**	1	.348**	.651**	.647**	.579**
	Sig.	.079	.000		.002	.000	.000	.000
	N	75	75	75	75	75	75	75
Belief Conflict	P- Corr	.371**	.971**	.348**	1	.330**	.366**	.269**
	Sig.	.001	.000	.002		.004	.000	.000
	N	75	75	75	75	75	75	75
Logic	P- Corr	.177	.457**	.651**	.330**	1	.950**	.959**
	Sig.	.128	.000	.000	.004		.000	.000
	N	75	75	75	75	75	75	75
Logic No-conflict	P- Corr	.171	.488**	.647**	.366**	.950**	1	.823**
	Sig.	.141	.000	.000	.000	.000		.000
	N	75	75	75	75	75	75	75
Logic Conflict	P- Corr	.167	.390**	.579**	.269*	.959**	.823**	1
	Sig.	.152	.001	.000	.000	.000	.000	
	N	75	75	75	75	75	75	75

Key: P-Corr = Pearson Correlation. Sig. = 2-tailed. ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

Table C6.26 Correlations Matrix for Stroop Interference Effect (SIE) on Accuracy scores – N = 75

		SIE	Belief	Belief No-conflict	Belief Conflict	Logic	Logic No-conflict	Logic Conflict
SIE	P- Corr	1	-.330**	-.196	-.295	-.028	.043	-.090
	Sig.		.007	.092	.010	.814	.711	.444
	N	75	75	75	75	75	75	75
Belief	P- Corr	-.310**	1	.563**	.971**	.457**	.488**	.390**
	Sig.	.007		.000	.000	.000	.000	.001
	N	75	75	75	75	75	75	75
Belief No-conflict	P- Corr	-.196	.563**	1	.348**	.651**	.647**	.579**
	Sig.	.092	.000		.002	.000	.000	.000
	N	75	75	75	75	75	75	75
Belief Conflict	P- Corr	-.295*	.971**	.348**	1	.330**	.366**	.269*
	Sig.	.010	.000	.002		.004	.000	.000
	N	75	75	75	75	75	75	75
Logic	P- Corr	-.028	.457**	.651**	.330**	1	.950**	.959**
	Sig.	.814	.000	.000	.004		.000	.000
	N	75	75	75	75	75	75	75
Logic No-conflict	P- Corr	.043	.488**	.647**	.366**	.950**	1	.823**
	Sig.	.711	.000	.000	.000	.000		.000
	N	75	75	75	75	75	75	75
Logic Conflict	P- Corr	-.090	.390**	.579**	.269*	.959**	.823**	1
	Sig.	.444	.001	.000	.020	.000	.000	
	N	75	75	75	75	75	75	75

Key: P-Corr = Pearson Correlation. Sig. = 2-tailed. ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

Experiment 8

Accuracy

Table C7.1 Repeated measures ANOVA on N-Back Instruction only - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.172	1	.172	3.827	.054	.051
Problem Type x Condition	.136	1	.136	3.023	.086	.041
Error (Problem Type)	3.186	71	.045			
Complexity (MP/Disjunctives)	1.000	1	1.000	12.712	.001	.152
Complexity x Condition	.043	1	.043	.541	.465	.008
Error (Complexity)	5.587	71	.079			
Problem x Complexity	7.383E-005	1	7.383E-005	.002	.963	.000
Problem x Complexity x Condition	.033	1	.033	.985	.324	.014
Error (Problem * Complexity)	2.412	71	.034			

Note: Condition is a between subjects factor

Table C7.2 Between subject effects on N-Back Instruction only - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Intercept	378.623	1	378.623	1194.037	.000	.944
Condition	5.433	1	5.433	17.132	.000	.194
Error	22.514	71	.317			

Table C7.3 Repeated measures ANOVA on N-Back Instruction only - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.039	1	.039	3.244	.076	.040
Problem Type x Condition	.005	1	.005	.438	.510	.006
Error (Problem Type)	.926	77	.012			
Complexity (MP/Disjunctives)	.247	1	.247	11.015	.001	.125
Complexity x Condition	.016	1	.016	.725	.397	.009
Error (Complexity)	1.729	77	.022			
Problem x Complexity	.016	1	.016	2.019	.159	.026
Problem x Complexity x Condition	.003	1	.003	.415	.521	.005
Error (Problem * Complexity)	.615	77	.008			

Note: Condition is a between subjects factor

Table C7.4 Between subject effects on N-Back Instruction only - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Intercept	225.477	1	225.477	2833.439	.000	.974
Condition	.642	1	.642	8.064	.006	.095
Error	6.127	77	.080			

Table C7.5 Repeated measures ANOVA on Belief & Logic Instruction - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.634	1	1.634	22.482	.000	.240
Instruction x Condition (Experimental or Control)	.038	1	.038	.518	.474	.007
Error (Instruction)	5.161	71	.073			
Problem Type (Conflict/No-conflict)	1.917	1	1.917	37.058	.000	.343
Problem Type x Condition	.001	1	.001	.014	.906	.000
Error (Problem Type)	3.673	71	.052			
Complexity (MP/Disjunctives)	2.412	1	2.412	37.944	.000	.348
Complexity x Condition	.132	1	.132	2.082	.153	.028
Error (Complexity)	4.512	71	.064			
Instruction x Problem Type	.788	1	.788	16.034	.000	.184
Instruction x Problem Type x Condition	.444	1	.444	9.035	.004	.113
Error (Instruction * Problem Type)	3.488	71	.049			
Instruction x Complexity	1.066	1	1.066	27.773	.000	.281
Instruction x Complexity x Condition	.003	1	.003	.073	.788	.001
Error (Instruction * Complexity)	2.725	71	.038			
Problem x Complexity	.482	1	.482	10.255	.002	.126
Problem x Complexity x Condition	.146	1	.146	3.108	.082	.042
Error (Problem * Complexity)	3.339	71	.047			
Instruction x Problem x Complexity	.077	1	.077	1.698	.197	.023
Instruction x Problem x Complexity x Condition	5.421E-006	1	5.421E-006	.000	.991	.000
Error (Instruction x Problem x Complexity)	3.219	71	.045			

Note: Condition is a between subjects factor

Table C7.6 Between subject effects on Belief & Logic Instruction - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Intercept	938.704	1	938.704	1826.430	.000	.963
Condition	.467	1	.467	.909	.344	.013
Error	36.491	71	.514			

Table C7.7 Repeated measures ANOVA on Belief & Logic Instruction - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.455	1	.455	16.324	.000	.175
Instruction x Condition (Experimental or Control)	.007	1	.007	.263	.610	.003
Error (Instruction)	2.148	77	.028			
Problem Type (Conflict/No-conflict)	1.128	1	1.128	27.301	.000	.262
Problem Type x Condition	.001	1	.001	.023	.880	.000
Error (Problem Type)	3.180	77	.041			
Complexity (MP/Disjunctives)	.371	1	.371	30.455	.000	.283
Complexity x Condition	.000	1	.000	.033	.856	.000
Error (Complexity)	.939	77	.012			
Instruction x Problem Type	.463	1	.463	19.132	.000	.199
Instruction x Problem Type x Condition	.026	1	.026	1.060	.306	.014
Error (Instruction * Problem Type)	1.865	77	.024			
Instruction x Complexity	.176	1	.176	19.407	.000	.201
Instruction x Complexity x Condition	.000	1	.000	.016	.899	.000
Error (Instruction * Complexity)	.700	77	.009			
Problem x Complexity	.131	1	.131	13.443	.000	.149
Problem x Complexity x Condition	.019	1	.019	1.905	.172	.024
Error (Problem * Complexity)	.751	77	.010			
Instruction x Problem x Complexity	.060	1	.060	5.468	.022	.066
Instruction x Problem x Complexity x Condition	.002	1	.002	.139	.710	.002
Error (Instruction x Problem x Complexity)	.850	77	.011			

Note: Condition is a between subjects factor

Table C7.8 Between subject effects on Belief & Logic Instruction - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Intercept	488.694	1	488.694	3325.712	.000	.977
Condition	.107	1	.107	.728	.396	.009
Error	11.315	77	.147			

Follow Up Analysis

Table C7.9 Repeated measures ANOVA on Experimental Condition - excluded data set N = 37

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.099	1	1.099	16.268	.000	.311
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	2.432	36	.068			
Problem Type (Conflict/No-conflict)	.934	1	.934	18.142	.000	.335
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	1.854	36	.052			
Complexity (MP/Disjunctives)	1.862	1	1.862	26.784	.000	.427
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	2.503	36	.070			
Instruction x Problem Type	1.224	1	1.224	18.727	.000	.342
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	2.352	36	.065			
Instruction x Complexity	.486	1	.486	15.983	.000	.307
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	1.096	36	.030			
Problem x Complexity	.588	1	.588	10.592	.002	.227
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.998	36	.055			
Instruction x Problem x Complexity	.040	1	.040	.835	.367	.023
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	1.712	36	.048			

Table C7.10 Repeated measures ANOVA on Control Condition - excluded data set N = 36

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.580	1	.580	7.438	.010	.175
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	2.729	35	.078			
Problem Type (Conflict/No-conflict)	.983	1	.983	18.911	.000	.351
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	1.819	35	.052			
Complexity (MP/Disjunctives)	.697	1	.697	12.149	.001	.258
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	2.009	35	.057			
Instruction x Problem Type	.024	1	.024	.745	.394	.021
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	1.135	35	.032			
Instruction x Complexity	.581	1	.581	12.480	.001	.263
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	1.629	35	.047			
Problem x Complexity	.048	1	.048	1.254	.270	.035
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.341	35	.038			
Instruction x Problem x Complexity	.037	1	.037	.867	.358	.024
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	1.507	35	.043			

Latencies

Table C7.11 Repeated measures ANOVA on N-Back Instruction only - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	89762.405	1	89762.405	.144	.705	.002
Problem Type x Condition	1378778.651	1	1378778.651	2.217	.141	.030
Error (Problem Type)	44148983.931	71	621816.675			
Complexity (MP/Disjunctives)	14058837.595	1	14058837.595	22.106	.000	.237
Complexity x Condition	3137907.650	1	3137907.650	4.934	.030	.065
Error (Complexity)	45154165.932	71	635974.168			
Problem x Complexity	733558.090	1	733558.090	1.431	.236	.020
Problem x Complexity x Condition	16324.665	1	16324.665	.032	.859	.000
Error (Problem * Complexity)	36406310.917	71	512764.942			

Note: Condition is a between subjects factor

Table C7.12 Between subject effects on N-Back Instruction only - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Intercept	4150416200.125	1	4150416200.125	722.995	.000	.911
Condition	78591821.495	1	78591821.495	13.691	.000	.162
Error	407581977.498	71	5740591.232			

Table C7.13 Repeated measures ANOVA on N-Back Instruction only - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	12141.641	1	12141.641	.013	.909	.000
Problem Type x Condition	2283788.729	1	2283788.729	2.483	.119	.031
Error (Problem Type)	70811497.068	77	919629.832			
Complexity (MP/Disjunctives)	22955792.423	1	22955792.423	20.385	.000	.209
Complexity x Condition	5920963.132	1	5920963.132	5.258	.025	.064
Error (Complexity)	86708559.551	77	1126085.189			
Problem x Complexity	1090932.508	1	1090932.508	1.112	.295	.014
Problem x Complexity x Condition	829.292	1	829.292	.001	.977	.000
Error (Problem * Complexity)	75541349.391	77	981056.486			

Note: Condition is a between subjects factor

Table C7.14 Between subject effects on N-Back Instruction only - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Intercept	4958630414.642	1	4958630414.642	388.110	.000	.834
Condition	36473822.642	1	36473822.642	2.855	.095	.036
Error	983779176.801	77	12776352.945			

Table C7.15 Repeated measures ANOVA on Belief & Logic Instruction - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1516913.785	1	1516913.785	.796	.375	.011
Instruction x Condition (Experimental or Control)	7530276.484	1	7530276.484	3.949	.051	.053
Error (Instruction)	135372483.756	71	1906654.701			
Problem Type (Conflict/No-conflict)	14900245.775	1	14900245.775	9.097	.004	.114
Problem Type x Condition	1035446.460	1	1035446.460	.632	.429	.009
Error (Problem Type)	116294430.383	71	1637949.724			
Complexity (MP/Disjunctives)	9473566.228	1	9473566.228	6.009	.017	.078
Complexity x Condition	3634590.543	1	3634590.543	2.305	.133	.031
Error (Complexity)	111935912.833	71	1576562.153			
Instruction x Problem Type	1578455.037	1	1578455.037	1.608	.209	.022
Instruction x Problem Type x Condition	59414.325	1	59414.325	.061	.806	.001
Error (Instruction * Problem Type)	69695723.648	71	981629.911			
Instruction x Complexity	16009295.293	1	16009295.293	18.414	.000	.206
Instruction x Complexity x Condition	377755.937	1	377755.937	.434	.512	.006
Error (Instruction * Complexity)	61728763.583	71	869419.205			
Problem x Complexity	3639745.552	1	3639745.552	4.161	.045	.055
Problem x Complexity x Condition	2116743.552	1	2116743.552	2.420	.124	.033
Error (Problem * Complexity)	62108171.941	71	874762.985			
Instruction x Problem x Complexity	142204.834	1	142204.834	.114	.737	.002
Instruction x Problem x Complexity x Condition	4785.190	1	4785.190	.004	.951	.000
Error (Instruction x Problem x Complexity)	88898741.830	71	1252094.955			

Note: Condition is a between subjects factor

Table C7.16 Between subject effects on Belief & Logic Instruction - excluded data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Intercept	22941223818.108	1	22941223818.108	1420.085	.000	.952
Condition	19663628.834	1	19663628.834	1.217	.274	.017
Error	1146992207.385	71	16154819.822			

Table C7.17 Repeated measures ANOVA on Belief & Logic Instruction - full data set – N = 79

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	186269.202	1	186269.202	.044	.834	.001
Instruction x Condition (Experimental or Control)	2977456.544	1	2977456.544	.706	.403	.009
Error (Instruction)	324826640.045	77	4218527.793			
Problem Type (Conflict/No-conflict)	34211613.113	1	34211613.113	13.165	.001	.146
Problem Type x Condition	3040086.379	1	3040086.379	1.170	.283	.015
Error (Problem Type)	200096457.336	77	2598655.290			
Complexity (MP/Disjunctives)	16482027.653	1	16482027.653	7.433	.008	.088
Complexity x Condition	3028921.526	1	3028921.526	1.366	.246	.017
Error (Complexity)	170738268.835	77	2217380.115			
Instruction x Problem Type	11304224.937	1	11304224.937	3.176	.079	.040
Instruction x Problem Type x Condition	685236.785	1	685236.785	.193	.662	.002
Error (Instruction * Problem Type)	274026435.936	77	3558784.882			
Instruction x Complexity	11039175.692	1	11039175.692	5.824	.018	.070
Instruction x Complexity x Condition	15917023.490	1	15917023.490	8.397	.005	.098
Error (Instruction * Complexity)	145950841.396	77	1895465.473			
Problem x Complexity	2547517.134	1	2547517.134	1.331	.252	.017
Problem x Complexity x Condition	1152104.375	1	1152104.375	.602	.440	.008
Error (Problem * Complexity)	147395974.359	77	1914233.433			
Instruction x Problem x Complexity	532710.952	1	532710.952	.198	.658	.003
Instruction x Problem x Complexity x Condition	30819.306	1	30819.306	.011	.915	.000
Error (Instruction x Problem x Complexity)	207568508.067	77	2695694.910			

Note: Condition is a between subjects factor

Table C7.18 Between subject effects on Belief & Logic Instruction - full data set – N = 79

	SS	df	MS	F	p	η_p^2
Intercept	27130119637.403	1	27130119637.403	815.504	.000	.914
Condition	42761810.567	1	42761810.567	1.285	.260	.016
Error	2561630268.268	77	33267925.562			

Follow Up Analysis

Table C7.19 Repeated measures ANOVA on Experimental Condition - excluded data set N = 37

	SS	df	MS	F	p	η_p^2
Instruction (Belief/Logic)	8013124.338	1	8013124.338	4.478	.041	.111
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	64421413.912	36	1789483.720			
Problem Type (Conflict/No-conflict)	4096055.405	1	4096055.405	2.911	.097	.075
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	50656529.845	36	1407125.829			
Complexity (MP/Disjunctives)	695684.122	1	695684.122	.530	.471	.015
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	47222996.628	36	1311749.906			
Instruction x Problem Type	1140801.946	1	1140801.946	.970	.331	.026
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	42330767.804	36	1175854.661			
Instruction x Complexity	10800668.122	1	10800668.122	11.366	.002	.240
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	34210036.128	36	950278.781			
Problem x Complexity	103987.514	1	103987.514	.132	.718	.004
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	28268297.736	36	785230.493			
Instruction x Problem x Complexity	48067.514	1	48067.514	.033	.856	.001
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	52098221.236	36	1447172.812			

Table C7.20 Repeated measures ANOVA on Control Condition - excluded data set N = 38

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1128377.531	1	1128377.531	.557	.461	.016
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	70951069.844	35	2027173.424			
Problem Type (Conflict/No-conflict)	11734993.837	1	11734993.837	6.257	.017	.152
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	65637900.538	35	1875368.587			
Complexity (MP/Disjunctives)	12254137.670	1	12254137.670	6.628	.014	.159
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	64712916.205	35	1848940.463			
Instruction x Problem Type	505766.531	1	505766.531	.647	.427	.018
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	27364955.844	35	781855.881			
Instruction x Complexity	5656846.420	1	5656846.420	7.195	.011	.171
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	27518727.455	35	786249.356			
Problem x Complexity	5577521.670	1	5577521.670	5.769	.022	.141
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	33839874.205	35	966853.549			
Instruction x Problem x Complexity	98235.281	1	98235.281	.093	.762	.003
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	36800520.594	35	1051443.446			

Individual Differences

Table C7.21 Repeated measures ANOVA on excluded accuracy data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.526	1	1.526	21.191	.000	.230
Instruction x CRT (high or low group)	.085	1	.085	1.182	.281	.016
Error (Instruction)	5.114	71	.072			
Problem Type (Conflict/No-conflict)	1.751	1	1.751	35.462	.000	.333
Problem Type x CRT	.168	1	.168	3.393	.070	.046
Error (Problem Type)	3.506	71	.049			
Complexity (MP/Disjunctives)	2.293	1	2.293	35.563	.000	.334
Complexity x CRT	.067	1	.067	1.040	.311	.014
Error (Complexity)	4.578	71	.064			
Instruction x Problem Type	.785	1	.785	14.176	.000	.166
Instruction x Problem Type x CRT	.001	1	.001	.019	.891	.000
Error (Instruction * Problem Type)	3.930	71	.055			
Instruction x Complexity	1.023	1	1.023	26.731	.000	.274
Instruction x Complexity x CRT	.010	1	.010	.265	.608	.004
Error (Instruction * Complexity)	2.717	71	.038			
Problem x Complexity	.511	1	.511	10.503	.002	.129
Problem x Complexity x CRT	.028	1	.028	.577	.450	.008
Error (Problem * Complexity)	3.457	71	.049			
Instruction x Problem x Complexity	.064	1	.064	1.424	.237	.020
Instruction x Problem x Complexity x CRT	.034	1	.034	.756	.388	.011
Error (Instruction x Problem x Complexity)	3.185	71	.045			

Note: CRT group (high and low) as a between subjects factor

Table C7.22 Between subject effects on excluded accuracy data set – N = 73

	SS	df	MS	F	*p*	η_p^2
Intercept	936.812	1	936.812	1924.282	.000	.964
CRT group	2.393	1	2.393	4.914	.030	.065
Error	34.565	71	.487			

Table C7.23 Correlations Matrix for Global Span (GS) scores against both Instruction and Problem Types – N = 73

		Global Span Score	Belief	Belief No-conflict	Belief Conflict	Logic	Logic No-conflict	Logic Conflict
GS Scores	P- Corr	1	.135	.290*	.046	.187	.212	.151
	Sig.		.253	.013	.697	.112	.072	.202
	N	73	73	73	73	73	73	73
Belief	P- Corr	.135	1	.833**	.965**	.691**	.746**	.587**
	Sig.	.253		.000	.000	.000	.000	.000
	N	73	73	73	73	73	73	73
Belief No-conflict	P- Corr	.290*	.833**	1	.659**	.743**	.796**	.635**
	Sig.	.013	.000		.000	.000	.000	.000
	N	73	73	73	73	73	73	73
Belief Conflict	P- Corr	.046	.965**	.659**	1	.588**	.637**	.497**
	Sig.	.697	.000	.000		.000	.000	.000
	N	73	73	73	73	73	73	73
Logic	P- Corr	.187	.691**	.743**	.588**	1	.962**	.962**
	Sig.	.112	.000	.000	.000		.000	.000
	N	73	73	73	73	73	73	73
Logic No-conflict	P- Corr	.212	.746**	.796**	.637**	.575**	1	.800**
	Sig.	.072	.000	.000	.000	.000		.000
	N	73	73	73	73	73	73	73
Logic Conflict	P- Corr	.151	.587**	.635**	.497**	.962**	.800**	1
	Sig.	.202	.000	.000	.000	.000	.000	
	N	73	73	73	73	73	73	73

Key: P-Corr = Pearson Correlation. Sig. = 2-tailed. ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

Table C7.24 Correlations Matrix for Stroop Interference Effect (SIE) on Accuracy scores – N = 73

		SIE	Belief	Belief No-conflict	Belief Conflict	Logic	Logic No-conflict	Logic Conflict
SIE	P- Corr	1	.206	.228	.171	.228	.210	.222
	Sig.		.081	.052	.147	.052	.074	.059
	N	73	73	73	73	73	73	73
Belief	P- Corr	.206	1	.833**	.965**	.691**	.746**	.587**
	Sig.	.081		.000	.000	.000	.000	.001
	N	73	73	73	73	73	73	73
Belief No-conflict	P- Corr	.228	.833**	1	.659**	.743**	.796**	.635**
	Sig.	.052	.000		.000	.000	.000	.000
	N	73	73	73	73	73	73	73
Belief Conflict	P- Corr	.171	.965**	.659**	1	.588**	.497**	.269*
	Sig.	.147	.000	.000		.000	.020	.000
	N	73	73	73	73	73	73	73
Logic	P- Corr	.228	.691**	.743**	.588**	1	.934**	.962**
	Sig.	.052	.000	.000	.000		.000	.000
	N	73	73	73	73	73	73	73
Logic No-conflict	P- Corr	.210	.746**	.796**	.637**	.934**	1	.800**
	Sig.	.074	.000	.000	.000	.000		.000
	N	73	73	73	73	73	73	73
Logic Conflict	P- Corr	.222	.587**	.635**	.497**	.962**	.800**	1
	Sig.	.059	.001	.000	.020	.000	.000	
	N	73	73	73	73	73	73	73

Key: P-Corr = Pearson Correlation. Sig. = 2-tailed. ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

Experiment 9

Accuracy

Table C8.1 Repeated measures ANOVA on Stroop Instruction only - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	1.126	1	1.126	37.066	.000	.283
Problem Type x Condition	1.669	1	1.669	54.942	.000	.369
Error (Problem Type)	2.855	94	.030			
Complexity (MP/Disjunctives)	.458	1	.458	34.873	.000	.271
Complexity x Condition	1.522	1	1.522	115.982	.000	.552
Error (Complexity)	1.234	94	.013			
Problem x Complexity	.364	1	.364	21.957	.000	.189
Problem x Complexity x Condition	1.156	1	1.156	69.692	.000	.426
Error (Problem * Complexity)	1.559	94	.017			

Note: Condition is a between subjects factor

Table C8.2 Between subject effects on Stroop Instruction only - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Intercept	798.615	1	798.615	14568.755	.000	.994
Condition	.320	1	.320	5.845	.018	.059
Error	5.153	94	.055			

Table C8.3 Repeated measures ANOVA on Stroop Instruction only - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.082	1	.082	29.409	.000	.211
Problem Type x Condition	.110	1	.110	39.395	.000	.264
Error (Problem Type)	.306	110	.003			
Complexity (MP/Disjunctives)	.017	1	.017	4.875	.029	.042
Complexity x Condition	.141	1	.141	41.072	.000	.272
Error (Complexity)	.377	110	.003			
Problem x Complexity	.027	1	.027	16.635	.000	.131
Problem x Complexity x Condition	.077	1	.077	46.915	.000	.299
Error (Problem * Complexity)	.180	110	.002			

Note: Condition is a between subjects factor

Table C8.4 Between subject effects on Stroop Instruction only - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Intercept	410.598	1	410.598	18972.745	.000	.994
Condition	.024	1	.024	1.120	.292	.010
Error	2.381	110	.022			

<u>Follow Up Analysis on Stroop Instruction</u>

Table C8.5 Repeated measures ANOVA on Experimental Condition - excluded data set N = 50

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	2.888	1	2.888	93.371	.000	.656
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	1.516	49	.031			
Complexity (MP/Disjunctives)	1.904	1	1.904	215.481	.000	.815
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	.433	49	.009			
Problem x Complexity	1.470	1	1.470	106.408	.000	.685
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	.677	49	.014			

Table C8.6 Repeated measures ANOVA on Control Condition - excluded data set N = 46

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	.026	1	.026	.859	.359	.019
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	1.339	45	.030			
Complexity (MP/Disjunctives)	.149	1	.149	8.376	.006	.157
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	.801	45	.018			
Problem x Complexity	.107	1	.107	5.447	.024	.108
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	.882	45	.020			

Analysis on Belief & Logic Instruction

Table C8.7 Repeated measures ANOVA on Belief & Logic Instruction - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.046	1	2.046	22.637	.000	.194
Instruction x Condition (Experimental or Control)	.213	1	.213	2.354	.128	.024
Error (Instruction)	8.497	94	.090			
Problem Type (Conflict/No-conflict)	4.808	1	4.808	56.043	.000	.374
Problem Type x Condition	.013	1	.013	.153	.697	.002
Error (Problem Type)	8.065	94	.086			
Complexity (MP/Disjunctives)	4.633	1	4.633	50.773	.000	.351
Complexity x Condition	.125	1	.125	1.368	.245	.014
Error (Complexity)	8.577	94	.091			
Instruction x Problem Type	.401	1	.401	8.476	.004	.083
Instruction x Problem Type x Condition	.175	1	.175	3.710	.057	.038
Error (Instruction * Problem Type)	4.446	94	.047			
Instruction x Complexity	.087	1	.087	1.571	.213	.016
Instruction x Complexity x Condition	.045	1	.045	.809	.371	.009
Error (Instruction * Complexity)	5.216	94	.055			
Problem x Complexity	.491	1	.491	17.322	.000	.156
Problem x Complexity x Condition	.054	1	.054	1.917	.169	.020
Error (Problem * Complexity)	2.666	94	.028			
Instruction x Problem x Complexity	.182	1	.182	6.131	.015	.061
Instruction x Problem x Complexity x Condition	.025	1	.025	.856	.357	.009
Error (Instruction x Problem x Complexity)	2.798	94	.030			

Note: Condition is a between subjects factor

Table C8.8 Between subject effects on Belief & Logic Instruction - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Intercept	1245.583	1	1245.583	3477.233	.000	.974
Condition	.004	1	.004	.010	.919	.000
Error	33.672	94	.358			

Table C8.9 Repeated measures ANOVA on Belief & Logic Instruction - full data set – N =112

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.827	1	.827	21.567	.000	.164
Instruction x Condition (Experimental or Control)	8.053E-006	1	8.053E-006	.000	.988	.000
Error (Instruction)	4.216	110	.038			
Problem Type (Conflict/No-conflict)	2.981	1	2.981	47.948	.000	.304
Problem Type x Condition	.293	1	.293	4.714	.032	.041
Error (Problem Type)	6.839	110	.062			
Complexity (MP/Disjunctives)	.639	1	.639	45.203	.000	.291
Complexity x Condition	.010	1	.010	.692	.407	.006
Error (Complexity)	1.555	110	.014			
Instruction x Problem Type	.739	1	.739	22.076	.000	.167
Instruction x Problem Type x Condition	.008	1	.008	.230	.632	.002
Error (Instruction * Problem Type)	3.681	110	.033			
Instruction x Complexity	.041	1	.041	4.536	.035	.040
Instruction x Complexity x Condition	.001	1	.001	.105	.746	.001
Error (Instruction * Complexity)	.996	110	.009			
Problem x Complexity	.154	1	.154	26.670	.000	.195
Problem x Complexity x Condition	.000	1	.000	.049	.825	.000
Error (Problem * Complexity)	.637	110	.006			
Instruction x Problem x Complexity	.030	1	.030	7.080	.009	.060
Instruction x Problem x Complexity x Condition	.000	1	.000	.092	.762	.001
Error (Instruction x Problem x Complexity)	.471	110	.004			

Note: Condition is a between subjects factor

Table C8.10 Between subject effects on Belief & Logic Instruction - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Intercept	671.995	1	671.995	4926.406	.000	.978
Condition	.445	1	.445	3.265	.074	.029
Error	15.005	110	.136			

Follow Up Analysis

Table C8.11 Repeated measures ANOVA on Experimental Condition - excluded data set N = 50

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.867	1	1.867	15.097	.000	.236
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	6.060	49	.124			
Problem Type (Conflict/No-conflict)	2.254	1	2.254	22.657	.000	.316
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	4.874	49	.099			
Complexity (MP/Disjunctives)	1.689	1	1.689	18.279	.000	.272
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	4.527	49	.092			
Instruction x Problem Type	.577	1	.577	8.871	.004	.153
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	3.190	49	.065			
Instruction x Complexity	.004	1	.004	.062	.805	.001
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	2.874	49	.059			
Problem x Complexity	.455	1	.455	17.277	.000	.261
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.291	49	.026			
Instruction x Problem x Complexity	.037	1	.037	.996	.323	.020
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	1.837	49	.037			

Table C8.12 Repeated measures ANOVA on Control Condition - excluded data set N = 46

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.451	1	.451	8.325	.006	.156
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	2.437	45	.054			
Problem Type (Conflict/No-conflict)	2.555	1	2.555	36.035	.000	.445
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	3.191	45	.071			
Complexity (MP/Disjunctives)	3.014	1	3.014	33.483	.000	.427
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	4.051	45	.090			
Instruction x Problem Type	.022	1	.022	.789	.379	.017
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	1.256	45	.028			
Instruction x Complexity	.123	1	.123	2.372	.131	.050
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	2.341	45	.052			
Problem x Complexity	.105	1	.105	3.437	.070	.071
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.375	45	.031			
Instruction x Problem x Complexity	.165	1	.165	7.742	.008	.147
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	.961	45	.021			

Table C8.13 Repeated measures ANOVA on Belief Instruction - excluded data set N = 96

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	3.993	1	3.993	41.622	.000	.307
Problem Type x Condition	.046	1	.046	.483	.489	.005
Error (Problem Type)	9.018	94	.096			
Complexity (MP/Disjunctives)	1.724	1	1.724	38.912	.000	.293
Complexity x Condition	.010	1	.010	.226	.636	.002
Error (Complexity)	4.166	94	.044			
Problem x Complexity	.636	1	.636	19.695	.000	.173
Problem x Complexity x Condition	.003	1	.003	.084	.773	.001
Error (Problem * Complexity)	3.037	94	.032			

Table C8.14 Repeated measures ANOVA on Logic Instruction - excluded data set N = 96

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	1.216	1	1.216	32.732	.000	.258
Problem Type x Condition	.142	1	.142	3.827	.053	.039
Error (Problem Type)	3.493	94	.037			
Complexity (MP/Disjunctives)	2.995	1	2.995	29.249	.000	.237
Complexity x Condition	.160	1	.160	1.560	.215	.016
Error (Complexity)	9.627	94	.102			
Problem x Complexity	.037	1	.037	1.451	.231	.015
Problem x Complexity x Condition	.077	1	.077	2.988	.087	.031
Error (Problem * Complexity)	2.427	94	.026			

Latencies

Table C8.15 Repeated measures ANOVA on Stroop Instruction only - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	4983111.086	1	4983111.086	7.145	.009	.071
Problem Type x Condition	2429471.638	1	2429471.638	3.483	.065	.036
Error (Problem Type)	65561594.820	94	697463.775			
Complexity (MP/Disjunctives)	2087359.223	1	2087359.223	5.400	.022	.054
Complexity x Condition	16762.963	1	16762.963	.043	.835	.000
Error (Complexity)	36332181.777	94	386512.572			
Problem x Complexity	6861.060	1	6861.060	.020	.888	.000
Problem x Complexity x Condition	248565.987	1	248565.987	.729	.396	.008
Error (Problem * Complexity)	32069274.846	94	341162.498			

Note: Condition is a between subjects factor

Table C8.16 Between subject effects on Stroop Instruction only - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Intercept	4908146569.626	1	4908146569.626	1203.040	.000	.928
Condition	110713088.074	1	110713088.074	27.137	.000	.224
Error	383499842.832	94	4079785.562			

Table C8.17 Repeated measures ANOVA on Stroop Instruction only - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Problem Type (Conflict/No-conflict)	7768622.021	1	7768622.021	10.948	.001	.091
Problem Type x Condition	3030427.736	1	3030427.736	4.271	.041	.037
Error (Problem Type)	78055648.693	110	709596.806			
Complexity (MP/Disjunctives)	821266.799	1	821266.799	1.320	.253	.012
Complexity x Condition	12512.871	1	12512.871	.020	.887	.000
Error (Complexity)	68423461.192	110	622031.465			
Problem x Complexity	259175.951	1	259175.951	.606	.438	.005
Problem x Complexity x Condition	330731.059	1	330731.059	.774	.381	.007
Error (Problem * Complexity)	47019961.049	110	427454.191			

Note: Condition is a between subjects factor

Table C8.18 Between subject effects on Stroop Instruction only - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Intercept	6538189397.755	1	6538189397.755	768.706	.000	.875
Condition	189878467.755	1	189878467.755	22.324	.000	.169
Error	935599156.165	110	8505446.874			

Table C8.19 Repeated measures ANOVA on Belief & Logic Instruction - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	3013043.349	1	3013043.349	2.150	.146	.022
Instruction x Condition (Experimental or Control)	1611165.224	1	1611165.224	1.150	.286	.012
Error (Instruction)	131706215.681	94	1401129.954			
Problem Type (Conflict/No-conflict)	55997624.262	1	55997624.262	34.661	.000	.269
Problem Type x Condition	4125699.158	1	4125699.158	2.554	.113	.026
Error (Problem Type)	151866042.081	94	1615596.192			
Complexity (MP/Disjunctives)	45654413.564	1	45654413.564	38.171	.000	.289
Complexity x Condition	7570312.991	1	7570312.991	6.329	.014	.063
Error (Complexity)	112430020.279	94	1196064.046			
Instruction x Problem Type	1487491.416	1	1487491.416	1.890	.172	.020
Instruction x Problem Type x Condition	5737682.031	1	5737682.031	7.290	.008	.072
Error (Instruction * Problem Type)	73980055.270	94	787021.865			
Instruction x Complexity	121314.182	1	121314.182	.121	.729	.001
Instruction x Complexity x Condition	710000.370	1	710000.370	.707	.402	.007
Error (Instruction * Complexity)	94346101.379	94	1003681.930			
Problem x Complexity	2333474.918	1	2333474.918	1.835	.179	.019
Problem x Complexity x Condition	433471.751	1	433471.751	.341	.561	.004
Error (Problem * Complexity)	119550380.175	94	1271812.555			
Instruction x Problem x Complexity	646307.002	1	646307.002	.556	.458	.006
Instruction x Problem x Complexity x Condition	311.950	1	311.950	.000	.987	.000
Error (Instruction x Problem x Complexity)	109261188.799	94	1162353.072			

Note: Condition is a between subjects factor

Table C8.20 Between subject effects on Belief & Logic Instruction - excluded data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Intercept	26412220278.153	1	26412220278.153	1293.496	.000	.932
Condition	7340356.372	1	7340356.372	.359	.550	.004
Error	1919409811.273	94	20419253.311			

Table C8.21 Repeated measures ANOVA on Belief & Logic Instruction - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	349780.034	1	349780.034	.062	.804	.001
Instruction x Condition (Experimental or Control)	795568.034	1	795568.034	.141	.708	.001
Error (Instruction)	621982940.750	110	5654390.370			
Problem Type (Conflict/No-conflict)	75877965.051	1	75877965.051	15.626	.000	.124
Problem Type x Condition	5902653.810	1	5902653.810	1.216	.273	.011
Error (Problem Type)	534141147.002	110	4855828.609			
Complexity (MP/Disjunctives)	72152633.368	1	72152633.368	14.552	.000	.117
Complexity x Condition	2385261.180	1	2385261.180	.481	.489	.004
Error (Complexity)	545417270.185	110	4958338.820			
Instruction x Problem Type	9259183.374	1	9259183.374	2.080	.152	.019
Instruction x Problem Type x Condition	7108742.392	1	7108742.392	1.597	.209	.014
Error (Instruction * Problem Type)	489727502.741	110	4452068.207			
Instruction x Complexity	3904199.830	1	3904199.830	1.344	.249	.012
Instruction x Complexity x Condition	1854155.401	1	1854155.401	.638	.426	.006
Error (Instruction * Complexity)	319527653.848	110	2904796.853			
Problem x Complexity	4214352.266	1	4214352.266	1.438	.233	.013
Problem x Complexity x Condition	171492.400	1	171492.400	.059	.809	.001
Error (Problem * Complexity)	322270435.849	110	2929731.235			
Instruction x Problem x Complexity	3688530.556	1	3688530.556	1.171	.282	.011
Instruction x Problem x Complexity x Condition	846929.931	1	846929.931	.269	.605	.002
Error (Instruction x Problem x Complexity)	346488707.747	110	3149897.343			

Note: Condition is a between subjects factor

Table C8.22 Between subject effects on Belief & Logic Instruction - full data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Intercept	34373588941.000	1	34373588941.000	811.515	.000	.881
Condition	24363686.687	1	24363686.687	.575	.450	.005
Error	4659304674.178	110	42357315.220			

Follow Up Analysis

Table C8.23 Repeated measures ANOVA on Experimental Condition - excluded data set N = 50

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	4711721.422	1	4711721.422	5.469	.023	.100
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	42212709.953	49	861483.877			
Problem Type (Conflict/No-conflict)	47229194.522	1	47229194.522	24.168	.000	.330
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	95754420.353	49	1954171.844			
Complexity (MP/Disjunctives)	8370316.922	1	8370316.922	8.817	.005	.153
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	46516391.452	49	949314.111			
Instruction x Problem Type	6818104.323	1	6818104.323	9.586	.003	.164
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	34851785.052	49	711260.919			
Instruction x Complexity	127484.702	1	127484.702	.120	.731	.002
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	52234077.173	49	1066001.575			
Problem x Complexity	2493083.103	1	2493083.103	2.600	.113	.050
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	46981248.272	49	958800.985			
Instruction x Problem x Complexity	352182.903	1	352182.903	.388	.536	.008
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	44510061.973	49	908368.612			

Table C8.24 Repeated measures ANOVA on Control Condition - excluded data set N = 46

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	104456.522	1	104456.522	.053	.820	.001
Instruction x Condition (Experimental or Control)	.000	0	.	.	.	.000
Error (Instruction)	89493505.728	45	1988744.572			
Problem Type (Conflict/No-conflict)	14267531.522	1	14267531.522	11.442	.001	.203
Problem Type x Condition	.000	0	.	.	.	.000
Error (Problem Type)	56111621.728	45	1246924.927			
Complexity (MP/Disjunctives)	43395045.924	1	43395045.924	29.626	.000	.397
Complexity x Condition	.000	0	.	.	.	.000
Error (Complexity)	65913628.826	45	1464747.307			
Instruction x Problem Type	663510.533	1	663510.533	.763	.387	.017
Instruction x Problem Type x Condition	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	39128270.217	45	869517.116			
Instruction x Complexity	680776.043	1	680776.043	.727	.398	.016
Instruction x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction * Complexity)	42112024.207	45	935822.760			
Problem x Complexity	362632.348	1	362632.348	.225	.638	.005
Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Problem * Complexity)	72569131.902	45	1612647.376			
Instruction x Problem x Complexity	296745.924	1	296745.924	.206	.652	.005
Instruction x Problem x Complexity x Condition	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	64751126.826	45	1438913.929			

Individual Differences

Table C8.25 Repeated measures ANOVA on excluded accuracy data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	1.973	1	1.973	22.551	.000	.193
Instruction x CRT (high or low group)	.488	1	.488	5.577	.020	.056
Error (Instruction)	8.222	94	.087			
Problem Type (Conflict/No-conflict)	4.608	1	4.608	56.335	.000	.375
Problem Type x CRT	.389	1	.389	4.753	.032	.048
Error (Problem Type)	7.689	94	.082			
Complexity (MP/Disjunctives)	4.479	1	4.479	48.900	.000	.342
Complexity x CRT	.091	1	.091	.994	.321	.010
Error (Complexity)	8.611	94	.092			
Instruction x Problem Type	.389	1	.389	8.225	.005	.080
Instruction x Problem Type x CRT	.174	1	.174	3.670	.058	.038
Error (Instruction * Problem Type)	4.448	94	.047			
Instruction x Complexity	.082	1	.082	1.467	.229	.015
Instruction x Complexity x CRT	5.019E-005	1	5.019E-005	.001	.976	.000
Error (Instruction * Complexity)	5.260	94	.056			
Problem x Complexity	.464	1	.464	17.364	.000	.156
Problem x Complexity x CRT	.206	1	.206	7.719	.007	.076
Error (Problem * Complexity)	2.514	94	.027			
Instruction x Problem x Complexity	.173	1	.173	5.780	.018	.058
Instruction x Problem x Complexity x CRT	.003	1	.003	.114	.736	.001
Error (Instruction x Problem x Complexity)	2.820	94	.030			

Note: CRT group (high and low) as a between subjects factor

Table C8.26 Between subject effects on excluded accuracy data set – N = 96

	SS	df	MS	F	*p*	η_p^2
Intercept	1249.946	1	1249.946	3793.036	.000	.976
CRT group	2.699	1	2.699	8.191	.005	.080
Error	30.976	94	.330			

Table C8.27 Repeated measures ANOVA on full accuracy data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.633	1	.633	18.296	.000	.143
Instruction x CRT (high or low group)	.408	1	.408	11.784	.001	.097
Error (Instruction)	3.808	110	.035			
Problem Type (Conflict/No-conflict)	2.603	1	2.603	43.668	.000	.284
Problem Type x CRT	.576	1	.576	9.660	.002	.081
Error (Problem Type)	6.556	110	.060			
Complexity (MP/Disjunctives)	.590	1	.590	42.255	.000	.278
Complexity x CRT	.029	1	.029	2.048	.155	.018
Error (Complexity)	1.537	110	.014			
Instruction x Problem Type	.536	1	.536	18.447	.000	.144
Instruction x Problem Type x CRT	.494	1	.494	16.992	.000	.134
Error (Instruction * Problem Type)	3.195	110	.029			
Instruction x Complexity	.034	1	.034	3.765	.055	.033
Instruction x Complexity x CRT	.013	1	.013	1.472	.228	.013
Error (Instruction * Complexity)	.984	110	.009			
Problem x Complexity	.117	1	.117	23.099	.000	.174
Problem x Complexity x CRT	.080	1	.080	15.824	.000	.126
Error (Problem * Complexity)	.557	110	.005			
Instruction x Problem x Complexity	.027	1	.027	6.422	.013	.055
Instruction x Problem x Complexity x CRT	.002	1	.002	.531	.468	.005
Error (Instruction x Problem x Complexity)	.469	110	.004			

Note: CRT group (high and low) as a between subjects factor

Table C8.28 Between subject effects on full accuracy data set – N = 112

	SS	df	MS	F	*p*	η_p^2
Intercept	666.517	1	666.517	5467.966	.000	.980
CRT group	2.042	1	2.042	16.749	.000	.132
Error	13.408	110	.122			

Table C8.29 Repeated measures ANOVA on High CRT group only - excluded accuracy data set – N = 45

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	.235	1	.235	4.967	.031	.101
Instruction x CRT (high or low group)	.000	0	.	.	.	.000
Error (Instruction)	2.078	44	.047			
Problem Type (Conflict/No-conflict)	1.092	1	1.092	16.065	.000	.267
Problem Type x CRT	.000	0	.	.	.	.000
Error (Problem Type)	2.990	44	.068			
Complexity (MP/Disjunctives)	1.550	1	1.550	21.948	.000	.333
Complexity x CRT	.000	0	.	.	.	.000
Error (Complexity)	3.107	44	.071			
Instruction x Problem Type	.020	1	.020	.493	.486	.011
Instruction x Problem Type x CRT	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	1.802	44	.041			
Instruction x Complexity	.041	1	.041	.689	.411	.015
Instruction x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction * Complexity)	2.592	44	.059			
Problem x Complexity	.024	1	.024	.977	.328	.022
Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.092	44	.025			
Instruction x Problem x Complexity	.060	1	.060	2.178	.147	.047
Instruction x Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	1.218	44	.028			

Table C8.30 Repeated measures ANOVA on Low CRT group only - excluded accuracy data set – N = 51

	SS	df	MS	F	*p*	η_p^2
Instruction (Belief/Logic)	2.359	1	2.359	19.192	.000	.277
Instruction x CRT (high or low group)	.000	0	.	.	.	.000
Error (Instruction)	6.144	50	.123			
Problem Type (Conflict/No-conflict)	4.093	1	4.093	43.549	.000	.466
Problem Type x CRT	.000	0	.	.	.	.000
Error (Problem Type)	4.699	50	.094			
Complexity (MP/Disjunctives)	3.119	1	3.119	28.332	.000	.362
Complexity x CRT	.000	0	.	.	.	.000
Error (Complexity)	5.504	50	.110			
Instruction x Problem Type	.577	1	.577	10.912	.002	.179
Instruction x Problem Type x CRT	.000	0	.	.	.	.000
Error (Instruction * Problem Type)	2.646	50	.053			
Instruction x Complexity	.042	1	.042	.780	.381	.015
Instruction x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction * Complexity)	2.668	50	.053			
Problem x Complexity	.688	1	.688	24.204	.000	.326
Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Problem * Complexity)	1.421	50	.028			
Instruction x Problem x Complexity	.120	1	.120	3.754	.058	.070
Instruction x Problem x Complexity x CRT	.000	0	.	.	.	.000
Error (Instruction x Problem x Complexity)	1.602	50	.032			

Table C8.31 Correlations Matrix for Global Span (GS) scores against both Instruction and Problem Types – N = 96

		Global Span Score	Belief	Belief No-conflict	Belief Conflict	Logic	Logic No-conflict	Logic Conflict
GS Scores	P-Corr	1	.268**	.178	.256*	.115	.212*	.028
	Sig.		.008	.083	.012	.266	.038	.787
	N	96	96	96	96	96	96	96
Belief	P-Corr	.268**	1	.640**	.961**	.471**	.359**	.483**
	Sig.	.008		.000	.000	.000	.000	.000
	N	96	96	96	96	96	96	96
Belief No-conflict	P-Corr	.178	.640**	1	.404**	.603**	.597**	.519**
	Sig.	.083	.000		.000	.000	.000	.000
	N	96	96	96	96	96	96	96
Belief Conflict	P-Corr	.256*	.961**	.404**	1	.345**	.214*	.389**
	Sig.	.012	.000	.000		.001	.036	.000
	N	96	96	96	96	96	96	96
Logic	P-Corr	.115	.471**	.603**	.345**	1	.880**	.940**
	Sig.	.266	.000	.000	.001		.000	.000
	N	96	96	96	96	96	96	96
Logic No-conflict	P-Corr	.212*	.359**	.597**	.214*	.880**	1	.666**
	Sig.	.038	.000	.000	.036	.000		.000
	N	96	96	96	96	96	96	96
Logic Conflict	P-Corr	.028	.483**	.519**	.389**	.940**	.666**	1
	Sig.	.787	.000	.000	.000	.000	.000	
	N	96	96	96	96	96	96	96

Key: P-Corr = Pearson Correlation. Sig. = 2-tailed. ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.

Table C8.32 Correlations Matrix for Stroop Interference Effect (SIE) on Accuracy scores – N = 96

		SIE	Belief	Belief No-conflict	Belief Conflict	Logic	Logic No-conflict	Logic Conflict
SIE	P-Corr	1	-.244*	-.009	-.288**	-.103	.033	-.185
	Sig.		.016	.929	.004	.317	.753	.071
	N	96	96	96	96	96	96	96
Belief	P-Corr	-.244*	1	.640**	.961**	.471**	.359**	.483**
	Sig.	.016		.000	.000	.000	.000	.000
	N	96	96	96	96	96	96	96
Belief No-conflict	P-Corr	-.009	.640**	1	.404**	.603**	.597**	.519**
	Sig.	.929	.000		.000	.000	.000	.000
	N	96	96	96	96	96	96	96
Belief Conflict	P-Corr	-.288**	.961**	.404**	1	.345**	.214*	.389**
	Sig.	.004	.000	.000		.001	.036	.000
	N	96	96	96	96	96	96	96
Logic	P-Corr	-.103	.471**	.603**	.345**	1	.880**	.940**
	Sig.	.317	.000	.000	.001		.000	.000
	N	96	96	96	96	96	96	96
Logic No-conflict	P-Corr	.033	.359**	.597**	.214*	.880**	1	.666**
	Sig.	.753	.000	.000	.036	.000		.000
	N	96	96	96	96	96	96	96
Logic Conflict	P-Corr	-.185	.483**	.519**	.389**	.940**	.666**	1
	Sig.	.071	.000	.000	.000	.000	.000	
	N	96	96	96	96	96	96	96

Key: P-Corr = Pearson Correlation. Sig. = 2-tailed. ** = Correlation is significant at the 0.01 level (2-tailed): * = 0.05 level.