

# Examination of the Belief Bias Effect across Two Domains of Reasoning

by

Nadia Martin

A thesis  
presented to the University of Waterloo  
in fulfillment of the  
thesis requirement for the degree of  
Master of Arts  
in  
Psychology

Waterloo, Ontario, Canada, 2008

©Nadia Martin 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

The belief bias effect – the finding that prior beliefs influence judgments of logic and evidence – has been a topic of much empirical investigation in both deductive and causal reasoning. However, to date, no research has examined the degree to which such biases are the result of common or distinct mechanisms in these two domains. By using common scales of measurement, I examine the degree to which individuals show common biases in these two domains in two experiments. Surprisingly, although the belief bias effect was observed in both paradigms, biases in one domain were unreliably associated with biases in the other domain. Experiment 2 included 6 measures of individual differences in an attempt to uncover the observation of differential biases in these domains. Dogmatism was found to be the single most predictive measure of belief bias, but only in deductive reasoning. These data are discussed in terms of dual process theories of reasoning.

## Acknowledgements

I would like to thank Jonathan Fugelsang, Derek Koehler, Colin MacLeod, and the University of Waterloo, Department of Psychology, Cognition division for their support. I would also like to thank my three exceptional research assistants, Frank Preston, Billiejoe Rogers and Zhaleh Semnani-Azad, for their hard work and dedication.

## Dedication

This thesis is dedicated to Ned, who has supported me and who has believed in me since the beginning of my studies.

## Table of Contents

|  |      |
|--|------|
| Author's Declaration.....  | ii   |
| Abstract.....  | iii  |
| Acknowledgements.....  | iv   |
| Dedication.....  | v    |
| Table of Contents.....   | vi   |
| List of Figures.....   | viii |
| List of Tables.....  | ix   |
| Introduction.....  | 1    |
| Deductive Reasoning.....   | 2    |
| Causal Reasoning.....  | 4    |
| Dual-Process Theories.....   | 7    |
| Individual Differences.....  | 9    |
| Confidence Ratings.....  | 11   |
| The Current Experiments.....   | 12   |
| Experiment 1.....  | 13   |
| Method.....  | 14   |
| Participants.....  | 14   |
| Materials.....   | 14   |
| Design.....  | 16   |
| Procedure.....   | 16   |
| Results.....   | 17   |
| Acceptance Ratings.....  | 17   |
| Deductive Reasoning.....   | 18   |
| Causal Reasoning.....  | 18   |
| Confidence Ratings.....  | 19   |
| Beliefs and Logic Across Reasoning Domains - Correlational Analysis..... | 20   |
| Summary.....   | 21   |
| Experiment 2.....  | 23   |
| Method.....  | 23   |
| Participants.....  | 23   |
| Materials.....   | 24   |

|   |    |
|---|----|
| Reasoning Tasks .....   | 24 |
| Thinking Dispositions .....   | 24 |
| Cognitive Ability.....  | 26 |
| Design .....  | 27 |
| Procedure .....   | 27 |
| Results.....  | 29 |
| Acceptance Ratings.....   | 29 |
| Deductive Reasoning .....   | 30 |
| Causal Reasoning.....   | 32 |
| Confidence Ratings .....  | 33 |
| Deductive Reasoning .....   | 33 |
| Causal Reasoning.....   | 35 |
| Beliefs and Logic Across Reasoning Domains - Correlational Analysis.....      | 36 |
| Individual Differences.....   | 38 |
| Summary .....   | 38 |
| General Discussion .....  | 39 |
| Reasoning ability.....  | 40 |
| Finding an appropriate belief baseline: "Neutral" may not be so neutral ..... | 41 |
| Conflicts and confidence.....   | 41 |
| Defining belief .....   | 42 |
| Conclusion .....  | 43 |
| Appendix A Deductive Problems .....   | 44 |
| Appendix B Causal Scenarios.....  | 50 |
| Appendix C Thinking Dispositions.....   | 57 |
| Appendix D Cognitive Ability.....   | 61 |
| References.....   | 66 |

## List of Figures

|          |                           |   |
|----------|---------------------------|---|
| Figure 1 | Müller-Lyer Illusion..... | 9 |
|----------|---------------------------|---|

## List of Tables

|         |   |    |
|---------|---|----|
| Table 1 | Event frequencies used for the computation of the $\Delta P$ values used in Experiments 1 and 2.....  | 16 |
| Table 2 | Experiment 1: Acceptance Ratings.....   | 17 |
| Table 3 | Experiment 1: Confidence Ratings.....   | 19 |
| Table 4 | Correlations Between Indexes of the Use of Beliefs (Belief-D, Belief-C) and the Indexes of the Use of Logic (Logic-D, Logic-C) for each Task..... | 21 |
| Table 5 | Experiment 2: Acceptance Ratings.....   | 30 |
| Table 6 | Experiment 2: Confidence Ratings.....   | 34 |
| Table 7 | Correlations Between Indexes of Belief/Logic Use and Individual Differences..   | 37 |

## Introduction

In everyday life, professionals and laypeople alike are constantly exposed to abundant information, which they continuously absorb and evaluate in order to make choices and perform different activities. It is often assumed that when people's knowledge increases, they become better informed, in turn improving their performance and helping them make better choices. According to previous research, an accumulation of domain-specific knowledge helps inform and improve our decisions. Experts such as firefighters (Klein, 1993) or master chess players (Chase & Simon, 1973) for instance, have been shown to benefit from accuracy and speed advantage in conducting their tasks and in deciding on the best course of action to follow.

Although it is reasonable to assume that accumulation of knowledge can enhance performance and decision-making, this is not always the case. Several researchers have demonstrated that previous knowledge and beliefs can create a variety of biases. These biases reveal themselves in economic decisions (e.g., the framing effect: Tversky & Kahneman, 1981), in the evaluation that individuals make of other people (e.g., fundamental attribution error: Ross, 1977), and in the influence expectations and information can have on what is attended to (e.g., confirmation bias: Wason, 1960, 1966) and remembered (e.g., false memories: Loftus and Palmer, 1974; Loftus, 1975).

Prior beliefs can also influence reasoning. This is best observed when the conclusion of a logical problem contradicts one's expectations. The effect of prior beliefs on judgments of logic, termed the belief bias effect, has been studied extensively using the deductive reasoning paradigm (Evans, 2002). The influence of prior beliefs on reasoning has also been observed in causal reasoning (e.g., Fugelsang & Thompson, 2000, 2003; Fugelsang & Dunbar, 2005). Despite similar findings in these two areas of reasoning, the relation between the two domains has not yet been explored systematically within the same study. Therefore, some questions have been left

unanswered: Do the same individuals tend to be swayed by their beliefs across disparate domains of reasoning, or, do different people use their prior knowledge differently based on the task at hand? These questions cannot be answered readily given that past research has looked at these two tasks independently. Thus, the goal of the following experiments is to examine the extent to which common and distinct mechanisms underlie the preferential use of beliefs and logic across both deductive and causal reasoning tasks. Before laying out the details of my experiments, past studies and findings that guide the rationale and choices of the research paradigms will be examined.

### *Deductive Reasoning*

Deductive reasoning is a closed system of reasoning, with all necessary information given in the body of the argument. Syllogisms are a sub-type of deductive reasoning, being limited to two premises and one conclusion. Based on the premises given, the task is to evaluate the correctness (i.e., the logical validity) of the conclusion rather than the truth of the statement. A syllogism is considered valid when the conclusion follows the premises in a *necessary* manner. As such, if the premises are taken to be true, then the conclusion must be true. The validity of the syllogism is based on the logical structure of the problem rather than on the content of the premises and conclusion. Yet, the common finding is that people's propensity to accept a syllogism as valid or invalid is not solely a function of its logical form, but that, when it contains familiar content, it is also modulated by the believability of its conclusion (Evans, Barston & Pollard, 1983). The belief bias effect – the finding that of the acceptance of the conclusion as logical is influenced by prior beliefs – has traditionally been studied through the use of the deductive reasoning paradigm (Evans, 2002), which includes syllogistic reasoning. The belief bias effect predicts that if two syllogisms are presented, both containing a believable conclusion, it should be easier to accept the valid one as correct (e.g., All men are mortal; Socrates is a man;

Therefore, Socrates is mortal.) than to reject the invalid one as incorrect (e.g., All men are mortal; Socrates is mortal; Therefore, Socrates is a man.).

Wilkins (1929) was the first researcher to examine the influence of using familiar, rather than abstract, material on the performance on the syllogistic reasoning task. For example, familiar material such as nouns and adjectives would replace the letters in “Some A are C” to create content-laden statements such as “Some flowers are red”. At that time, Wilkins’ goal was to correlate success on various versions of the syllogistic reasoning task with participants’ scores on the Thorndike Intelligence Examination rather than a primary interest in the effect of prior beliefs on reasoning. This interest in correlating a measure of intelligence with performance on a syllogistic task reflected the long time view that deductive reasoning is an indicator of the highest level of reasoning ability and intellectual achievement. This view, which originated with ancient Greek philosophers such as Aristotle, is also present in Piaget’s theory of cognitive development (Inhelder & Piaget, 1958). According to Piaget, deductive reasoning ability is part of the formal operations stage. This stage of cognitive development involves the development of systematic thinking and the manipulation of abstractions. Reaching this level does not guarantee that a person will make constant use of logic, but rather that these operations are now accessible to her. In addition, some individuals will never attain this stage.

Nonetheless, this interest in correlating intelligence with deductive reasoning ability is not what set Wilkins’ research apart. Her true innovation was the use of syllogisms embedded in familiar material where the truth or falsity of the conclusion, in light of participant’s prior knowledge, was at variance with the logical status (i.e., valid or invalid) of the syllogism. For instance, the conclusion “Some cats are black”, despite being true in the realm of common knowledge, might follow invalidly from the premises. In contrast, the conclusion “All men are women”, which clearly contradicts general knowledge, can technically follow validly from a set

of premises. It is the use of such familiar material that allows the creation of conflict-laden reasoning problems (i.e., problems where the conclusion drawn based on the underlying logical structure is in opposition to the conclusion drawn in light of prior beliefs or knowledge). In fact, the inclusion of valid/non-believable and invalid/believable problems has been found repeatedly to give rise to what is now termed the belief bias effect.

At the heart of this effect are three robust findings (see Klauer, Musch, & Naumer, 2000 for a review): believable (i.e., knowledge-consistent) conclusions are accepted more often than non-believable (i.e., knowledge-inconsistent) conclusions; logically valid conclusions are accepted more often than logically invalid conclusions; finally, a believability by logic interaction appears due to the greater impact of validity in the non-believable condition, than the believable condition. To better understand the locus of this bias, a knowledge-independent (i.e., abstract or neutral) condition is sometimes added (e.g., Newstead, Pollard, Evans, Allen, 1992; Goel & Dolan, 2001; Thompson, 1996). This can be achieved either by using non-sense words or by replacing the words with letters. Under this condition, participants tend to evaluate neutral conclusions as they do believable ones. This is also found in other reasoning situations where beliefs and logic are in opposition (e.g., Klaczynski, Gordon & Fauth, 1997; Ditto, Munro, Apanovitch, Scepansky & Lockhart, 2003; Greenhoot, Semb, Colombo & Schreiber, 2004).

### *Causal Reasoning*

The well-defined characteristics of the syllogistic task, notably the existence of a clear normative standard (i.e., the validity of the conclusion being evaluated strictly based on its logical necessity from the premises), make it a paradigm of predilection to study the influence of prior beliefs on formal reasoning. Nonetheless, researchers have also studied the belief bias effect using causal reasoning tasks (e.g., Fugelsang & Thompson, 2001; Fugelsang & Dunbar, 2005). Normative standards in this reasoning domain can be, and are, hotly debated (e.g., White, 1988,

1989; Cheng, 1997), notably due to the probabilistic nature of the task. One of the key questions researchers have been concerned with centers around determining which cue best allows one to conclude that a causal link exists between two events. Although many cues to causation exist, such as generative transmission (Michotte, 1946, Shultz, 1982), mechanisms (Ahn, Kalish, Medin & Gelman, 1995), temporal contiguity (Wasserman & Neunaber, 1986; Young, 1995), spatial contiguity (Scholl & Tremoulet, 2000; Young, 1995) and covariation (Cheng & Novick, 1990; Shaklee & Elek, 1988), it is this last cue that has received the most empirical investigation, and is typically used as the normative standard by which to make a causality judgment (Rapus, 2001; Sà, Kelly, Ho & Stanovich, 2005; Cheng, 1997). In other words, the covariation between a putative cause and an effect, because it is based on facts or bottom-up information in opposition to prior beliefs, can be seen as the equivalent of the logical structure of the syllogism. This assumption is a guiding principle for the work contained in this thesis. For instance, a conclusion based on high covariation between two events would be taken as logical or valid, whereas a conclusion based on low covariation between two events would be seen as equivalent to an invalid problem structure. This view is supported by the empirical tradition of David Hume (1740/1938, 1739/1960) whereby observing some regularity in associations between events is what allows us to infer the cause of an effect. Similarly, Kelley (1967, 1973) emphasizes the role of covariation and contingency (i.e., the observation of the presence of the effect when the cause is present and the observation of the occurrence or non-occurrence of the effect when the cause is absent, in making causal attributions).

As noted previously, the difficulty in proposing a normative standard in causal reasoning is that, unlike deductive reasoning, the task is probabilistic in nature. Causal relations do not imply a constant pairing between a putative cause and the proposed effect. Also, observing the occurrence of event A immediately prior to event B does not guarantee that A is the cause of B.

Nonetheless, Cheng and Novick (1990), with their probabilistic contrast model, suggested a way to quantify the causal strength of an association normatively. To calculate this, the probability of obtaining an effect when the putative cause is present  $[P(e/c)]$  is compared to the probability of obtaining this same effect when the cause is absent  $[P(e/\sim c)]$  using the following unidirectional contingency rule:

$$\Delta P_c = [P(e/c)] - [P(e/\sim c)]$$

which allows the provision of some support for a facilitatory, inhibitory or neutral role of the putative cause.

Yet, many researchers disagree with a prevalent role of covariation in causal reasoning, either because they recognize other factors, such as those mentioned above, as having a greater impact on causal induction (e.g., White, 1988, 1989; Ahn et al, 1995) or because they have found people to be of limited ability when dealing with covariation information (e.g., Crocker, 1981; Smedlund, 1963). Others highlight the fact that anyone familiar with statistics should recognize that “correlation is not causation” (e.g., Koslowski, Okagaki, Lorenz & Umbach, 1989). This fact renders causal reasoning probabilistic rather than deterministic in nature, again highlighting an important difference between causal and deductive reasoning (although, see Evans, Handley, Harper & Johnson-Laird, 1999, regarding how participants still look at deductive problems with probability in mind). Therefore, participants are typically asked to respond on a multi-point scale to rate their degree of certitude that cause A is what led to effect B (i.e., using various Likert scales).

Despite the underlying differences between the tasks of those two domains of reasoning (i.e. deductive and causal), enough structural commonalities (i.e., an inherent logical structure and a possibility to vary the believability of the conclusion) exist to attempt a comparison between the two tasks. Indeed, the central question in this thesis is whether the same or different cognitive

mechanisms are involved in giving rise to this effect, both in deductive and in causal reasoning. We will now turn to a discussion of Dual-Process theories of reasoning, as they provide an important theoretical framework that can be used to generate hypotheses about how reasoning might proceed in these two disparate domains.

### *Dual-Process Theories*

The occurrence of belief bias effects in reasoning can be explained by dual-process theories (Evans, 2003, 2008). Many researchers have proposed different theories involving two “systems” (e.g., Sloman, 1996) or two “processes” operating in parallel (e.g., Epstein, 1994; Evans & Over, 1996). Stanovich and West (2000) have subsumed these different theories under the umbrella of System 1 and System 2. This neutral naming was intentional, underlying the fact that these classifications are debatable as they encompass many different concepts, some of which may be continuous rather than dichotomous in nature. Despite the possible limitations of this classification, this dichotomization is widely used, notably in explaining the occurrence of sub-optimal answers, judgments or decisions. System 1, a universally shared system (Evans, 2008) characterized as being fast and automatic, is often dubbed responsible for those sub-optimal performances. Due to its associative nature (Sloman, 1996), no calculations are necessary for this system to produce an answer. This reliance on associations explains its situational grounding. It is influenced by the current context, thus the appellation by Stanovich and West (1997, 1999, 2000) of the ability to think about a situation without the influence of the prior beliefs as the ability to “decontextualize”. This reliance on associations also allows more instinctual responses such as the use of heuristics. Evans (2003, 2008) has also qualified this system as implicit.

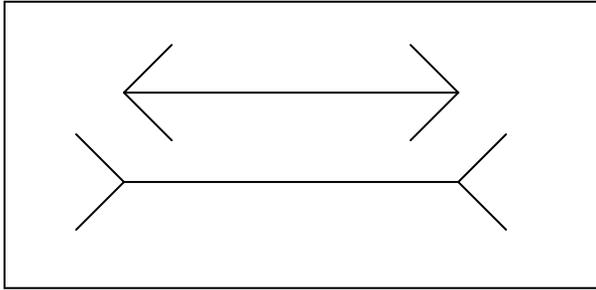
In contrast to System 1, System 2 is explicit, with the content of its operations being available to consciousness. Sequential and rule-based, System 2 has been defined as cognitively effortful and slow. It is assumed to underlie analytical and controlled processing, the type of

processing necessary to succeed on formal reasoning tasks for example. It has also been proposed that System 2 has the capacity to overrule System 1, but only when time, energy and motivation are available (Evans, 2003, 2008). In addition, given that System 2 is potentially related both to general intelligence and working memory capacity (Barrett, Tugade & Engle, 2004; Smith & DeCoster, 2000), only a sub-group of the population may be able to use System 2 in an efficient way. Thus, one would expect to find individual differences in the degree to which the operation of System 2 (and the interaction between System 1 and System 2 processes) unfold.

How can we infer the presence and interaction of the two systems in reasoning?

Appropriate tasks are those designed to put the two systems in conflict, forcing participants to respond with one system or the other. Sloman (1996) referred to those situations as meeting *Criterion S*. This criterion refers to the simultaneous emergence of two answers, one that refers to our first reaction (System 1) and a second that is the reflection of increased processing (System 2). This type of situation is often present when encountering perceptual illusions. For instance, when looking at a Müller-Lyer Illusion (see **Figure 1**), we are swayed by the apparent discrepancy between the two lines. This is taken to be a reflection of the work of System 1. However, we could decide to measure the lines instead of trusting our instinct, partly because we may recognize or know the potential influence of the inward- and outward-pointing arrows. Such a reflective process would be the result of the activation of System 2. If we think of Criterion S in light of the belief bias effect, we can see how a comparable situation can occur when beliefs and logic are in conflict. A response in line with prior beliefs would be considered as reflecting the activity of System 1. On the other hand, if System 2 succeeds in overriding System 1, then the answer should reflect the logic of the problem. Of course, only situations including a potential conflict between the two systems will allow the verification of this dual-process hypothesis as non-conflict situations should lead to the same answer regardless of which system is activated. As

mentioned earlier, individuals can be expected to differ in their use of each system when encountering conflicting or incongruent beliefs-logic situations. I am particularly interested in examining whether the same individuals will rely more prominently on their beliefs (i.e., System 1), or logic (i.e., System 2), in a consistent manner across both reasoning domains.



*Figure 1. Müller-Lyer Illusion*

#### *Individual Differences*

Research on individual differences in reasoning is the result of the convergence of two fields: the psychometric tradition and cognitive psychology. Whereas the psychometric tradition allows the use and the development of useful individual measures (e.g., I.Q. tests), cognitive psychology focuses on the processes of reasoning. Such research can possibly inform the rationality debate, helping to shed light on the appropriate normative standard, as well as possibly informing dual-process theories of cognition (Newstead, Handley, Harley, Wright & Farrelly, 2004).

With the work of Stanovich and West (e.g., 1997, 1998), individual differences are now receiving increased attention in the field of reasoning. First, these authors make a clear distinction between cognitive abilities and thinking dispositions. Cognitive abilities are related to how well we think, conduct computations and resolve problems. For example, the Backward Digit Span Task is often used to measure cognitive ability. Based on the WAIS-III (Wechsler, 1997), this test is set to measure working memory capacity (WMC). This is also used as a proxy for intelligence

(Smith & DeCoster, 2000), as a higher WMC score is thought to reflect greater cognitive ability. Alternatively, thinking dispositions are related to goal setting and motivation, therefore influencing the level of processing people will adopt when analyzing various types of information (Stanovich & West, 2000). The Cognitive Reflection Test (Frederick, 2005) is one recent addition to these measures. Here, Frederick (2005) has proposed a Cognitive Reflection Test to measure people's tendency to question their first intuitive answer before responding. A short questionnaire, each question is designed to quickly bring to mind an answer that can be modified if one takes time to think it through (e.g., "A bat and a ball cost \$1.10. The bat costs \$1.00 more than the ball. How much does the ball cost?"). Frederick relates a high score on this scale as an indication that one is more inclined to use System 2 over System 1.

These two aspects of reasoning (i.e., cognitive ability and thinking dispositions) are thought to be orthogonal. For example, an unmotivated student could do poorly in class despite a high level of cognitive ability. Both types of individual differences have been found to be predictive of reasoning ability, either when studied independently (e.g., Blais, Thompson & Baranski, 2005; Unsworth & Engle, 2006) or in concurrence (e.g., Stanovich & West, 1997; Sà et al., 2005; Macpherson & Stanovich, 2007). In the context of this study, I am interested in measuring both aspects as either or both can potentially inform the preferential use of beliefs or logic.

As we have seen so far, it is the combination of beliefs and logic into congruent and incongruent situations that lead to a variation in judgments. This will be examined in the current experiments using judgments of the logic of causal and deductive reasoning problems. The influence of combining beliefs and logic will also be examined using confidence ratings.

### *Confidence Ratings*

In studying the belief bias effect, the confidence participants have in their answers is rarely taken into consideration. When confidence ratings are included, it is typically under the assumption that people should display less confidence when they are wrong than when they are correct, thus demonstrating a certain level of metacognitive awareness. Shynkaruk and Thompson (2006) did study the relation between confidence and accuracy. They focused on syllogistic reasoning, systematically varying beliefs and logic to evoke the belief bias effect. In their experiment, participants gave two answers to each of 12 syllogistic reasoning problems. The first answer was set to be given 10 seconds after the beginning of the trial as an attempt to tap into the functioning of System 1 (Roberts & Newton, 2002; Evans & Curtis-Holmes, 2005). Based on dual-processing predictions, answers given quickly should reflect prior beliefs due to a lack of time to access System 2 analytical processing. Instructions also emphasized the need for participants to give the first answer that came to mind. A confidence rating followed this first answer. The second answer, set to be given within one minute of the first answer, was designed to allow enough time for System 2 to analyze and override the first answer, thus making sure that it would be in line with the normative logical standard (i.e., judging the validity of the problem without being swayed by the believability status of the conclusion). Even though participants tended to change their answer at the second opportunity when their self-reported confidence was low, this did not guarantee that the change was made toward the correct answer, but led to an increase of their second confidence rating. The authors explained this increase in confidence after the second answer as reflecting participants' rationalization of their answer rather than an analytical reconsideration of their first answer. Confidence ratings were also influenced by the believability status of the conclusion. When reasoning about a believable or unbelievable conclusion, in contrast to a neutral conclusion (e.g. "Some Abens are not Welps"), participants

felt more confident with their answers. Unlike their accuracy scores, this pattern in confidence ratings was similar for both good and poor reasoners. The authors concluded that various variables influence accuracy and confidence differently, and that self-reported confidence is not a good predictor of accuracy. In fact, one limitation of their analysis is their lack of focus on the interaction between beliefs and logic in influencing confidence ratings. I believe that it is the conflictive combinations that will lead to a decrease in confidence, independently of accuracy.

### *The Current Experiments*

The current experiments will attempt to address the issues raised in the introduction by examining participants responding (both in terms of accuracy and confidence) to both deductive and causal reasoning problems. Importantly, beliefs, and logic will be crossed in both domains, and participants will be exposed to all permutations of the problem types. This will allow a within-subject analysis of the commonalities between these two domains of reasoning. Whereas the first experiment will establish the methodology necessary to compare the two domains, the second experiment will in addition examine the question of individual differences in reasoning. The key questions addressed will be: (1) Do people respond with the same degree of bias to both deductive problems and causal problems? (2) How does the conflict between beliefs and logic influence participants' confidence in their answers? (3) Are there specific individual differences that may predict one's propensity to make use of beliefs and/or logic?

## Experiment 1

The first goal of this experiment was to explore the impact of the interaction between prior beliefs and logic, both in terms of acceptance ratings and confidence judgments in both tasks.

Specifically, it was expected that:

- a. Valid problems will be accepted as valid more often than invalid ones.
- b. Believable problems will be accepted as valid more often than non-believable ones.
- c. An interaction will occur between beliefs and logic. Based on past findings, the effect of believability should be greater for invalid problems in the deductive reasoning task. In contrast, the effect should be greater at the valid level of logic in the causal reasoning task. Thus, I predict a three-way interaction between, Beliefs, Logic, and Task.

In addition, the scores obtained by participants in judging the scenarios will be used to form indexes of belief use and logic use within each domain of reasoning. If it is assumed that humans have a general reasoning ability, then it should be found that the use of beliefs across tasks correlate positively. The same should be found in regard to the use of logic across tasks. If reasoning abilities are domain-specific, then no correlation should be found between the indexes across the two domains of reasoning. Independently of the generality or the specificity of reasoning ability, a negative correlation should be found between belief use and logic use in both reasoning tasks signifying that as one relies more on beliefs, they should rely less on logic, and vice versa.

The analysis of confidence ratings is novel. Instead of focusing on the influence of accuracy on confidence, the variable of interest is the presence or absence of a conflict between

beliefs and logic within the problem. It is assumed that conflictive problems should yield lower confidence ratings than problems presenting no conflict between beliefs and logic.

A secondary goal of this experiment is to test the validity of using a binary (i.e. yes/no scale) in both tasks. This will allow a direct comparison of performances on both the deductive and the causal task. Of course, this comparison was also made possible due to the use of a within-subject design, where each participant would complete each condition within each task.

## Method

### *Participants*

Sixty-four undergraduates ( $f = 43$ ,  $m = 21$ ) from the University of Waterloo participated for course credit. The mean age of this sample was 20.11 ( $SD = 3.322$ ) and the mean university level was 2.17 ( $SD = 1.115$ ).

### *Materials*

Eight syllogisms (**Appendix A**) and eight causal scenarios (**Appendix B**) were used in each of two versions of the materials booklets. The syllogisms were taken from the work of Evans et al. (1983). All were categorical syllogisms, composed of two premises followed by a conclusion. To avoid a possible confound due to a variation in the difficulty of the syllogisms, all problems followed the same EO-I format. Thus, the first premise, 'E', was of the form "No A is B" or "No B is A"; the second premise, 'O', was either of the form "Some C are B" or "Some B are C"; and finally, the conclusion, 'I', was always of the type "Some C are not A". The different combinations dictated the validity of the syllogisms, whereas the content of each conclusion dictated the believability status of the syllogism. The two versions of the deductive problems were achieved by reversing the validity status of the syllogism while keeping the believability constant. For example, the conclusion "Some priests are not religious" was used twice, once embedded in a valid syllogism and another time within an invalid syllogism. The causal scenarios

were adapted from the work of Fugelsang and Thompson (2003). As described previously for the syllogisms, reversing the validity status while keeping the believability constant created the two sets of causal reasoning problems. Each scenario debuted with a putative cause hypothesized to have led to a particular effect (see **Appendix B**). Covariation evidence was then provided that depicted the degree to which the candidate cause was statistically associated with the observed effect. This ‘evidence’ always concerned two samples sizes depicting the frequency of the occurrence and non-occurrence of the cause and effect relationship (see **Table 1** for the event frequencies used in the current experiment). The first event frequency provided information about the occurrence of the effect when the putative cause was present, whereas the second sample was always about the occurrence of the effect when the putative cause was absent. In an attempt to make both tasks as similar as possible, only two levels of covariation were used corresponding to  $\Delta P_c = 0$  and  $\Delta P_c = 1$ . These absolute values were chosen to mirror the valid and invalid structure of deductive reasoning problems respectively. The formula for  $\Delta P_c$  (Cheng & Novick, 1990), as described in the introduction, was used to derive the  $\Delta P_c$  scores.

Judgments of validity (for the syllogisms) (“Given that: ... Does it necessarily follow that: ...”) and causality (e.g., “Given the above information, do you think that insomnia caused the patients to be fatigued?”) were made using a simple binary (i.e., yes/no) response scale. These judgments were followed directly below by the question “On a scale of 0 to 100, how confident are you in your answer?” A space was provided for participants to write their confidence rating.

Table 1

Event frequencies used for the computation of the  $\Delta P$  values used in Experiments 1 and 2.

| Frequencies |     |     |      | $\Delta P$ Computations |   |         |            |
|-------------|-----|-----|------|-------------------------|---|---------|------------|
| ce          | c~e | ~ce | ~c~e | P(e/c)                  | - | P(e/~c) | $\Delta P$ |
| 10          | 0   | 0   | 10   | 10/10                   | - | 0/10    | 1          |
| 0           | 10  | 0   | 10   | 0/10                    | - | 0/10    | 0          |

Note. (ce) represents the number of times the cause and effect co-occurred; (c~e) represents the number of times the cause occurred in the absence of the effect; (~ce) represents the number of times the effect occurred in the absence of the cause.

### Design

A 2 (Task: deductive vs. causal) x 2 (Beliefs: believable vs. non-believable) x 2 (Logic: valid vs. invalid) within-subject design was used. The two dependent variables of interest were the acceptance ratings (i.e. judgments of validity/causality) and the confidence ratings. The order of the two tasks and their two versions was counterbalanced, yielding eight possible combinations. Within each task, the reasoning problems were presented in a random order.

### Procedure

After reading the information letter and signing the consent form, participants proceeded to complete both reasoning tasks. The instructions for each task (see **Appendices A and B**) underlined the importance of assuming the truthfulness of the given information, thus emphasizing the need to reason from logic rather than from prior beliefs. For each problem, participants first judged the validity of the conclusion on a yes/no scale. On the same page, participants also rated how confident they felt in their answer. When both tasks were completed,

participants had an opportunity to ask questions and received a feedback sheet. On average, the testing session lasted no longer than 20 minutes.

## Results

The results pertaining to each dependent variable - acceptance ratings and confidence ratings - will be reported separately. Nonetheless, the same analyses were conducted in each case. The results of the correlational analysis looking at the association between the use of beliefs and the use of logic across tasks are also reported.

### *Acceptance Ratings*

Data were analyzed using a 2 x 2 x 2 (Task x Beliefs x Logic) repeated-measures ANOVA. The mean acceptance ratings are presented in **Table 2**. A score of 1 represented the acceptance of a conclusion, whereas a score of 0 was associated with the rejection of a conclusion.

The three independent variables led to significant differences in acceptance ratings. As expected, believable conclusions were accepted more often than unbelievable conclusions,

*Table 2*

*Experiment 1: Acceptance Ratings*

| Logic          | Task      |      |        |      |
|----------------|-----------|------|--------|------|
|                | Deductive |      | Causal |      |
|                | M         | SD   | M      | SD   |
| <b>Valid</b>   |           |      |        |      |
| Believable     | .88       | .278 | .91    | .197 |
| Non-Believable | .50       | .418 | .66    | .407 |
| <b>Invalid</b> |           |      |        |      |
| Believable     | .65       | .375 | .12    | .278 |
| Non-Believable | .13       | .267 | .00    | .000 |

$F(1,63)=87.220$ ,  $MSE=.149$ ,  $p<.001$ , *partial*  $\eta^2=.581$ . The same was true for valid conclusions being accepted more often than invalid conclusions,  $F(1,63)=255.745$ ,  $MSE=.132$ ,  $p<.001$ , *partial*  $\eta^2=.802$ . As a first possible indication that people do not reason causally and deductively in the same manner, a main effect of Task was also found,  $F(1,63)=23.351$ ,  $MSE=.078$ ,  $p<.001$ , *partial*  $\eta^2=.270$ .

Several interactions were also found to be statistically significant. Whereas Beliefs and Logic did not interact when collapsed across tasks, Task did interact independently both with Beliefs,  $F(1,63)=29.180$ ,  $MSE=.080$ ,  $p<.001$ , *partial*  $\eta^2=.317$ , and Logic,  $F(1,63)=50.288$ ,  $MSE=.111$ ,  $p<.001$ , *partial*  $\eta^2=.444$ . More importantly, the three-way interaction was significant,  $F(1, 63) = 11.18$ ,  $MSE = .054$ ,  $p = .001$ , *partial*  $\eta^2 = .151$ . Because the three-way interaction was significant, two 2 x 2 (Beliefs x Logic) repeated measures ANOVAs were conducted; one for the deductive and one for the causal reasoning task.

*Deductive Reasoning.* In the deductive task, the typical belief bias effect was found, with a significant effect both of Beliefs,  $F(1,63)=85.045$ ,  $MSE=.155$ ,  $p<.001$ , *partial*  $\eta^2=.574$ , and Logic,  $F(1,63)=44.383$ ,  $MSE=.134$ ,  $p<.001$ , *partial*  $\eta^2=.413$ , as well as a significant interaction between Beliefs and Logic,  $F(1,63)=5.234$ ,  $MSE=.060$ ,  $p=.026$ , *partial*  $\eta^2=.077$ . This interaction was driven by a greater impact of Beliefs for the invalid problems ( $M = .52$ ,  $SD = .431$ ) than for the valid problems ( $M = .38$ ,  $SD = .494$ ), although beliefs had a significant effect for both invalid,  $t(63)=9.711$ ,  $p<.001$  and valid scenarios  $t(63)=6.200$ ,  $p<.001$  .

*Causal Reasoning.* Significant main effects of Beliefs,  $F(1,63)=29.194$ ,  $MSE=.074$ ,  $p<.001$ , *partial*  $\eta^2=.317$ , and of Logic,  $F(1,63)=305.624$ ,  $MSE=.109$ ,  $p<.001$ , *partial*  $\eta^2=.829$ , as well as the Beliefs by Logic interaction,  $F(1,63)=7.797$ ,  $MSE=.036$ ,  $p=.007$ , *partial*  $\eta^2=.110$ , were also found in the causal reasoning task. Here, in contrast to the deductive task, the interaction was

driven by a greater impact of Beliefs at the valid level of Logic ( $M = .25$ ,  $SD = .378$ ) in comparison to the invalid level ( $M = .12$ ,  $SD = .278$ ). Again, beliefs also had a significant effect for both valid,  $t(63)=5.292$ ,  $p<.001$ , and invalid scenarios  $t(63)=3.372$ ,  $p=.001$ ). This effect was simply larger for the valid scenarios.

*Confidence Ratings*

The mean confidence ratings are presented in **Table 3**. In a similar way that acceptance ratings vary when Beliefs and Logic are in conflict, confidence ratings decreased in both tasks under those incongruent conditions. Yet, the findings are different from those obtained with the acceptance ratings. A 2 x 2 x 2 (Task x Beliefs x Logic) repeated measures ANOVA was conducted. Neither the main effect of Task,  $F(1,63)=.506$ ,  $MSE=398.616$ ,  $p=.479$ , *partial*  $\eta^2 = .008$ , Beliefs,  $F(1,63)=.117$ ,  $MSE=174.654$ ,  $p=.733$ , *partial*  $\eta^2 = .002$ , nor Logic,  $F(1,63)=.028$ ,  $MSE=179.643$ ,  $p=.868$ , *partial*  $\eta^2 <.001$ , were significant. Both the Task by Beliefs interaction,

*Table 3*

*Experiment 1: Confidence Ratings*

| Logic          | Task      |        |        |        |
|----------------|-----------|--------|--------|--------|
|                | Deductive |        | Causal |        |
|                | M         | SD     | M      | SD     |
| Valid          |           |        |        |        |
| Believable     | 91.02     | 11.402 | 87.85  | 13.096 |
| Non-Believable | 84.30     | 18.172 | 78.99  | 16.940 |
| Invalid        |           |        |        |        |
| Believable     | 81.38     | 15.579 | 80.73  | 17.639 |
| Non-Believable | 87.59     | 14.532 | 91.70  | 12.151 |

$F(1,63)=.506$ ,  $MSE=398.616$ ,  $p=.479$ ,  $partial \eta^2 =.008$ , and the three-way interaction,  $F(1,63)=2.420$ ,  $MSE=157.353$ ,  $p=.125$ ,  $partial \eta^2 =.037$ , were also not significant, leaving only two of the two-way interactions to be significant. First, the Task by Logic interaction was significant,  $F(1,63)=6.937$ ,  $MSE=164.569$ ,  $p=.011$ ,  $partial \eta^2 =.099$ , which revealed that confidence ratings were affected by the logic of the problem only in the deductive task,  $F(1,63)=4.660$ ,  $MSE=.035$ ,  $p=.035$ ,  $partial \eta^2 =.069$ . The second significant interaction was the Beliefs by Logic interaction,  $F(1, 63) = 85.242$ ,  $MSE =100.703$ ,  $p < .001$ ,  $partial \eta^2 = .575$ . As predicted, this crossover interaction was caused by the confidence ratings in the conflict conditions being lower than in the no-conflict conditions.

#### *Beliefs and Logic Across Reasoning Domains – Correlational Analyses*

To allow an examination of participants' reasoning behaviour across tasks (i.e., to examine people's propensity to use beliefs versus logic within and across reasoning tasks), two types of difference scores were computed for each of the two tasks. The first one, a belief use index, involved adding the acceptance ratings for the believable conditions and subtracting the sum of the ratings from the non-believable conditions. Similarly, a logic use index was calculated by subtracting the sum of the acceptance ratings on invalid problems from the sum of the ratings on the valid problems. A correlation analysis was then conducted using these four difference scores (see **Table 4**). The use of logic was negatively correlated with the use of beliefs both in deductive reasoning,  $r(64) = -.479$ ,  $p < .001$ , and in causal reasoning,  $r(64) = -.804$ ,  $p < .001$ . Also, whereas a positive correlation was found between the use of beliefs across the tasks,  $r(64) = .323$ ,  $p < .01$ , the use of logic on one task was not predictive of using logic on the other task,  $r(64) = .086$ ,  $p > .05$ . This suggests that, whereas the use of beliefs might be more domain general, the use of logic might be a more domain specific skill. Thus, demonstrating good reasoning with

Table 4

*Correlations Between Indexes of the Use of Beliefs (Belief-D, Belief-C) and the Indexes of the Use of Logic (Logic-D, Logic-C) for each Task.*

| Index       | 1 | 2       | 3      | 4       |
|-------------|---|---------|--------|---------|
| 1. Belief-D |   | -.479** | .323** | -.147   |
| 2. Logic-D  |   |         | -.093  | -.086   |
| 3. Belief-C |   |         |        | -.804** |
| 4. Logic-C  |   |         |        |         |

\*\* . Correlation is significant at the .01 level (two-tailed).

one type of problem might not guarantee a similar ability when dealing with other types of reasoning tasks.

#### *Summary*

Main effects of Beliefs and Logic, as well as the interaction between these two factors were present in both tasks. Importantly, this effect was observed using a simple ‘yes/no’ scale common to the syllogistic task but not to causal reasoning. In line with previous research (e.g., Evans, Barston & Pollard, 1983; Fugelsang & Thompson, 2003), the pattern of interaction was different, however. Whereas beliefs had a greater influence on invalid problems in deductive reasoning, it is on valid problems that beliefs had the greatest impact in causal reasoning. It is encouraging to see that, despite using a much more restrictive yes/no scale, and using absolute covariation values (i.e.,  $\Delta P_c = 1$  and  $\Delta P_c = 0$ ), the belief bias effect was still obtained.

Indexes of belief use and logic use correlated negatively within each reasoning domain. The use of beliefs in the causal domain also correlated positively with the use of beliefs in deductive reasoning. In contrast, the use of logic on one task was not associated with the use of logic on the other task, defying the idea of a general reasoning ability. In addition, the results surrounding the confidence ratings made by participants supported the prediction that participants

are sensitive to conflicts between their prior beliefs and the logic of the problem. This sensitivity reduces participant's confidence in their response.

## Experiment 2

This experiment served as a replication and extension of Experiment 1. Specifically, it extended the findings of Experiment 1 in two important ways. First, I introduced a third level of beliefs. Whereas Experiment 1 only used believable and non-believable statements (based on common knowledge), this second experiment introduced neutral statements (simple letters were used) providing a neutral belief baseline and thus allowing a better understanding of the effect of beliefs on reasoning. I expected to replicate previous findings of a belief bias effect, with the acceptance ratings increasing in the believable/invalid condition and decreasing in the non-believable/valid condition. I also expected the confidence ratings to decrease again when beliefs and logic are in conflict. With regard to the neutral condition, I expected acceptance ratings to be similar to those in the believable condition.

Secondly, this experiment served to examine the degree to which reasoning biases in causal and deductive reasoning are associated with cognitive ability and thinking dispositions. I expected higher scores on the two cognitive ability tests and on three of the four included thinking dispositions measures (Need for Cognition, Preference for Numerical Information, and Cognitive Reflection Test) to correlate positively with the use of logic. Based on the results obtained in Experiment 1 regarding the lack of a correlation between the use of logic across the two tasks, I did not have a clear prediction on whether or not these measures would correlate with both or only one of the logic scores. In contrast, I expected higher dogmatism scores, the fourth thinking disposition measure, to correlate with a predominant use of beliefs in both tasks.

### Method

#### *Participants*

One hundred and five undergraduates ( $f = 57$ ,  $m = 48$ ) from the University of Waterloo participated either for course credit or for pay. The mean age was 21.51 ( $SD = 4.911$ ) and the mean university level was 2.57 ( $SD = 1.139$ ).

### *Materials*

Experiment 2 included eight tasks divided into three parts: two reasoning tasks, four measures of thinking dispositions and two measures of cognitive ability.

*Reasoning tasks.* As with Experiment 1, two reasoning tasks were administered: one deductive reasoning task and one causal reasoning task. The number of problems per task was increased in this second experiment due to the introduction of a third level of beliefs. Whereas Experiment 1 only used believable and non-believable statements and conclusion, this second experiment introduced neutral statements to allow a better understanding of the effect of beliefs on reasoning. Single letters were used instead of critical nouns and adjectives to create these neutral problems. For example, a syllogism would now appear as “No P is M; Some S are M; Therefore, Some S are not P”. The same technique was used to create neutral causal scenarios. All neutral problems can be found, by task, in **Appendices A and B**. As in Experiment 1, each condition within each task contained two problems, and two versions of each reasoning task were used. The two dependent variables used the same scales previously used in Experiment 1 (i.e., a ‘yes/no’ scale for the acceptance ratings and a ‘0 to 100’ scale for the confidence ratings).

*Thinking Dispositions.* Four different thinking dispositions scales were included: 1) Preference for Numerical Information Scale (PNI) (Viswanathan, 1993); 2) Need for Cognition Scale (NC) (Cacioppo, Petty & Kao, 1984); 3) Cognitive Reflection Test (CRT) (Frederick, 2005); and 4) Short Form Rokeach’s Dogmatism Scale (Dogma) (Troidahl & Powell, 1965) (See **Appendix C**).

1. *Preference for Numerical Information Scale (PNI)*: Designed to measure the proclivity to use numerical information, this scale was selected to potentially inform the lack of correlation of the two indexes of logic use. The scale includes 20 items such as “Understanding numbers is as important in daily life as reading or writing” or “I prefer not to pay attention to information involving numbers (reverse scored)”. Given that only the causal reasoning problems included numerical information, a correlation between PNI scores and the use of logic on the causal task, but not on the deductive task, could explain the aforementioned lack of correlation between the two domains of reasoning.
2. *Need for Cognition Scale (NC)*: One of the most widely used measures of thinking disposition in the literature is the Need for Cognition (Cacioppo, Petty & Kao, 1984). The scale includes 18 items such as “The notion of thinking abstractly is appealing to me” or “I only think as hard as I have to (reverse scored)”. Those high in NC are hypothesized to enjoy thinking and analytical activities to a greater extent than those low on that scale. Therefore, someone high on that scale would be expected to use logic over beliefs when confronted to conflicting situations.
3. *Cognitive Reflection Test (CRT)*: The CRT is hypothesized to measure people’s tendency to question their first intuitive answer, or not, before responding. Only three items compose this scale, which makes it a very attractive tool to use in research. All items are designed to easily bring to mind one answer, albeit the wrong one. For instance, when encountering this problem: “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?”, most people will automatically think “10 cents” – which is the wrong

answer. A high score on this scale is an indication that one is more likely to use System 2 to overcome the influence of System 1. It should therefore correlate positively with the use of logic on both tasks.

4. *Short Form Dogmatism Scale (Dogma)*: Dogmatism can be described as the tendency to hold on to one's own beliefs in face of contrary evidence. In the context of this research, it would relate to exclusive use of prior beliefs without consideration for the underlying logical structure. This short form version of the Dogmatism Scale includes 20 items (e.g., "There are two kinds of people in this world: those who are for the truth and those who are against the truth"; "It is often desirable to reserve judgment about what's going on until one has had a chance to hear the opinions of those one respects.") to which participants must indicate their degree of agreement. Given the importance given to their own beliefs and the discounting of evidence that characterize those high in dogmatism, a high score on this measure should correlate positively with the use of beliefs and negatively with the use of logic.

*Cognitive Ability*. Finally, the two measures of cognitive ability administered to participants were 1) Baddeley's 3 min reasoning test (Baddeley, 1968) and 2) a Backward Digit Span test (based on WAIS-III; Wechsler, 1997). (See **Appendix D**)

1. *Baddeley's 3 min reasoning test*: This cognitive ability task is a timed reasoning task. Participants have to indicate if two parts of a logical statement are congruent (e.g. 'A precedes B – AB' → true) or incongruent (e.g. 'A follows B – AB' → false). Thirty-two statements are included and participants have 90

seconds to complete the task<sup>1</sup>. This test was designed to measure people's reasoning ability. Scores on this test are expected to correlate positively with the use of logic on both reasoning tasks.

2. *Backward Digit Span test*: This test is designed to measure working memory capacity (WMC). A sequence of digits is said out loud and the participant's task is to repeat each sequence in reverse order. Each sequence includes between 2 and 9 digits. Each possible sequence length appears twice in a row (e.g., 4-5-8-2 and 3-4-9-7), with the length increasing by one digit after every second trial with the same sequence length. The task ends after a participant makes two subsequent mistakes. The highest sequence length successfully repeated represents the participant's score. Given that WMC is often found to correlate with general intelligence, it is also expected to be positively associated with the use of logic in both tasks.

### *Design*

A 2 (Task: deductive vs. causal) x 3 (Beliefs: believable vs. non-believable vs. neutral) x 2 (Logic: valid vs. invalid) within-subject design was used for the reasoning problems. As with Experiment 1, acceptance ratings and confidence ratings were the two dependent variables of interest in relation to the reasoning problems. Six measures of individual differences, which included four measures of thinking dispositions and two measures of cognitive ability, were also obtained from each participant (see *Materials*) for use in a correlational analysis.

### *Procedure*

---

<sup>1</sup> We elected to use 32 statements instead of the 64 proposed in the original article (Baddeley, 1968) due to an error in the latter in regard to the number of possible combinations. This also explains the use of a 90 seconds limit rather than the originally proposed 3 minutes.

Sessions were conducted with one or two participants in a room. After obtaining their written consent, each participant proceeded to complete both the causal and the deductive reasoning tasks. As with Experiment 1, participants accepted or rejected the conclusion using a yes/no scale for both types of reasoning problems. The problems within each task were randomized and presented in a booklet. The order of completion of the tasks and the versions of the tasks were counterbalanced.

Immediately after, participants completed a block of four thinking dispositions tests. The 4 tasks were counterbalanced across participants. Three out of the four measures, namely the PNI scale, the NC scale and the Dogmatism scale, were completed using a 7-point Likert-type scale (see **Appendix D**). For each of those three tasks, participants had to indicate their level of agreement with a series of statements. Some of the statements were reverse-scored. For the fourth task, the CRT, three short arithmetic problems embedded in a short text were presented in a constant order. Participants indicated their answers directly in the booklet.

The two cognitive ability tasks, also presented in a counterbalanced order, were always completed at the end of the session. This choice was made based on the need for the experimenter to interact with the participants in the completion of those tasks. The backward digit span task required the experimenter to say out loud a sequence of digits. Each sequence included between 2 and 9 digits. Each possible sequence length appeared twice in a row (e.g., 4-5-8-2 and 3-4-9-7) and the length was increased by one digit after every second trial with the same sequence length. Participants' task was to repeat each sequence in reverse order. The task ended after a participant made two subsequent mistakes. The highest sequence length successfully repeated was recorded as the participant's score. The second cognitive ability task was Baddeley's 3 min test (Baddeley, 1968), a timed reasoning task. Participants had to indicate if two parts of a logical statement were congruent (e.g. 'A precedes B – AB' → true) or incongruent (e.g. 'A follows B – AB' → false).

Thirty-two statements were included and participants had 90 seconds to complete the task. Finally, participants both received a feedback sheet and were given the opportunity to ask questions about the experiment.

## Results

As with the first experiment, repeated-measures ANOVAs were used to analyze both the acceptance ratings and the confidence ratings. We expected to replicate the pattern of results in regard to the both the acceptance and the confidence ratings as were obtained in Experiment 1 and extend the analyses to the neutral belief baseline in order to achieve a more precise measure of the effect of beliefs. Differences scores were again used to create separate indexes for each task with regard to the use of beliefs and the use of logic. If belief use is more general than logic use, then we expect to find the two indexes of belief use to correlate again across tasks. In addition to being correlated amongst themselves, these difference scores will also be correlated with the six measures of individual differences to examine the predictive nature of the cognitive ability and thinking dispositions.

### *Acceptance Ratings*

As a first step, we conducted a 2 x 3 x 2 (Task x Beliefs x Logic)<sup>2</sup> repeated measures ANOVA on the acceptance ratings, for which descriptive statistics are found in **Table 5**. The main effect of Task,  $F(1, 104)=69.115$ ,  $MSE=.196$ ,  $p<.001$ ,  $partial \eta^2= .399$ , Beliefs,  $F(2, 104)=71.453$ ,  $MSE=.129$ ,  $p<.001$ ,  $partial \eta^2=.407$ , and Logic,  $F(1,104)= 406.603$ ,  $MSE=.166$ ,  $p<.001$ ,  $partial \eta^2=.796$ , were all statistically significant. As previously obtained in Experiment 1, only two of the two-way interactions were significant. Whereas the Beliefs by Logic interaction failed to reach significance, both the Task by Beliefs,  $F(2, 208)=16.144$ ,  $MSE=123$ ,  $p<.001$ ,

---

<sup>2</sup> Despite a violation of sphericity (Mauchly's test: task x beliefs x logic,  $p<.05$ ), all corrections yielded highly significant results ( $p<.001$ ). For that reason, only values using pooled variance are reported.

Table 5

Experiment 2: Acceptance Ratings

| Logic          | Task      |      |        |      |
|----------------|-----------|------|--------|------|
|                | Deductive |      | Causal |      |
|                | M         | SD   | M      | SD   |
| <b>Valid</b>   |           |      |        |      |
| Believable     | .90       | .573 | .84    | .282 |
| Non-Believable | .55       | .392 | .67    | .372 |
| Neutral        | .80       | .338 | .90    | .281 |
| <b>Invalid</b> |           |      |        |      |
| Believable     | .69       | .388 | .12    | .275 |
| Non-Believable | .25       | .361 | .02    | .107 |
| Neutral        | .70       | .390 | .09    | .248 |

partial  $\eta^2=.134$ , and the Task by Logic,  $F(1,104)=146.836$ ,  $MSE=.146$ ,  $p<.001$ , partial  $\eta^2=.585$ , interactions were found to be statistically significant. More importantly, the 3-way interaction between Task, Beliefs and Logic was also significant,  $F(2,208)=12.340$ ,  $MSE=.076$ ,  $p<.001$ , partial  $\eta^2=.106$ . These different interactions supported our previous finding that people do not behave the same way across the two reasoning tasks.

For our next series of analysis, the data was split by Task to allow a better understanding of the influence of beliefs on logic within each domain of reasoning. The analyses related to each task will now be considered in succession.

*Deductive Reasoning.* For the deductive task, we first conducted a 3 x 2 (Beliefs x Logic) repeated measures ANOVA. The typical belief bias effect was found, with both the main effect of beliefs,  $F(2,208)=53.301$ ,  $MSE=.184$ ,  $p<.001$ , partial  $\eta^2= .339$ , and the main effect of Logic,

$F(1,104)=34.327$ ,  $MSE= .186$ ,  $p<.001$ ,  $partial \eta^2= .248$ , being significant. The typical Beliefs by Logic interaction,  $F(2,208)=6.255$ ,  $MSE=.097$ ,  $p=.002$ ,  $partial \eta^2=.057$ , was also present. To understand the origin of this interaction, we compared each level of Beliefs, two at a time, in a series of 2 x 2 (Beliefs x Logic) repeated measures ANOVA<sup>3</sup>. As predicted, when the Beliefs independent variable was restricted to only believable and neutral items, participants processed believable and neutral conclusions in a similar way. Indeed, the only significant finding in this analysis was the effect of Logic,  $F(1,104)=12.714$ ,  $MSE=.186$ ,  $p=.001$ ,  $partial \eta^2=.109$ , as no effect of Beliefs was found. Surprisingly, the comparison of acceptance ratings across the believable and non-believable conditions did not yield the usual interaction between Beliefs and Logic, leaving only the two main effects of Beliefs,  $F(1,104)=78.983$ ,  $MSE=.208$ ,  $p<.001$ ,  $partial \eta^2=.432$ , and Logic,  $F(1,104)=38.393$ ,  $MSE=.181$ ,  $p<.001$ ,  $\eta^2=.270$ . As will be discussed later, the addition of the neutral level of beliefs may have influenced the judgment of the other problems. A final 2 x 2 repeated measures ANOVA was conducted for the deductive task, comparing the non-believable to the neutral level of beliefs . As in the previous analysis, the two main effects of Beliefs,  $F(1,104)=88.007$ ,  $MSE=.146$ ,  $p<.001$ ,  $partial \eta^2=.458$ , and Logic,  $F(1,104)=39.812$ ,  $MSE=.103$ ,  $p<.001$ ,  $partial \eta^2=.277$ , were present. The interaction between Beliefs and Logic was also significant,  $F(1,104)=21.402$ ,  $MSE=.056$ ,  $p<.001$ ,  $partial \eta^2=.171$ . Post-hoc pairwise Bonferroni-corrected comparisons<sup>4</sup> revealed that the difference between these last two levels of beliefs was statistically significant both for the valid problems,  $t(104)= -5.466$ ,  $p<.001$ , and for the invalid problems,  $t(104)= -10.536$ ,  $p<.001$ , even though the impact of beliefs was greater at the invalid level of Logic (*Mean difference* =.46, *SD* =.44) than at the valid level of

---

<sup>3</sup> All 2 x 2 ANOVAs in Experiment 2 involved analyzing two levels of Beliefs at a time (Believable/Non-believable, Believe/Neutral, and Non-believable/Neutral), therefore excluding the third level of Beliefs in each of these analyses.

<sup>4</sup> For the two pairwise comparisons related to the deductive task, the Bonferroni-corrected  $\alpha=.025$ . For the four causal task pairwise comparisons, the Bonferroni-corrected  $\alpha=.0125$ .

Logic (*Mean difference* =.24, *SD* =.46). This difference in magnitude of the effects is in the same direction as reported in the past for the deductive task (e.g., Evans, 2002).

*Causal Reasoning.* The same analyses were conducted for the causal reasoning task as for the deductive reasoning task. For the 3 x 2 (Beliefs x Logic) repeated measures ANOVA, both Beliefs,  $F(2,208)=20.610$ ,  $MSE=.059$ ,  $p<.001$ , *partial*  $\eta^2=.165$ , and Logic,  $F(1,104)=656.627$ ,  $MSE=.098$ ,  $p<.001$ , *partial*  $\eta^2=.863$ , had a significant effect on the acceptance ratings, along with a significant interaction between these two variables,  $F(2,208)=6.905$ ,  $MSE=.054$ ,  $p=.001$ , *partial*  $\eta^2=.062$ . Because of this interaction, the three levels of Beliefs were compared two at a time using three 2 x 2 repeated measures ANOVA. First, our hypothesis regarding the similarity in the processing of believable and neutral conclusions was again supported, as the main effect of Beliefs was not significant for this analysis ( $F(1,104)=.3655$ ,  $MSE=.059$ ,  $p=.547$ , *partial*  $\eta^2=.003$ ). In contrast, the effect of Logic was highly significant,  $F(1,104)=623.148$ ,  $MSE=.098$ ,  $p<.001$ , *partial*  $\eta^2=.857$ . Surprisingly, an interaction was found between these two factors,  $F(1,104)=4.393$ ,  $MSE=.054$ ,  $p=.039$ , *partial*  $\eta^2=.041$ . However, post-hoc pairwise Bonferroni-corrected comparisons revealed no significant difference between the two types of beliefs, neither for the valid problems,  $t(104) = -1.920$ ,  $p=.058$ , nor for the invalid ones,  $t(104) = 1.000$ ,  $p=.320$ . In the second 2 x 2 repeated measures ANOVA, the believable and the non-believable levels were compared. As in the deductive task, the typical interaction was not present,  $F(1,104)= 2.768$ ,  $MSE=.117$ ,  $p=.099$ , *partial*  $\eta^2=.026$ . Only the two main effects of beliefs,  $F(1,104)= 25.432$ ,  $MSE=.073$ ,  $p<.001$ , *partial*  $\eta^2=.196$  and Logic,  $F(1,104)= 457.748$ ,  $MSE=.106$ ,  $p<.001$ , *partial*  $\eta^2=.815$ , were present. Finally, the comparison of non-believable to the neutral level of beliefs revealed significant main effects of beliefs,  $F(1,104)=32.434$ ,  $MSE=.071$ ,  $p<.001$ , *partial*  $\eta^2=.238$ , and Logic,  $F(1,104)= 571.689$ ,  $MSE=.097$ ,  $p<.001$ , *partial*  $\eta^2=.846$ . The Beliefs by Logic

interaction was also significant,  $F(1,104)= 12.584$ ,  $MSE=.055$ ,  $p=.001$ ,  $partial \eta^2= .108$ . Follow-up post-hoc pairwise Bonferroni-corrected comparisons revealed that the influence of Beliefs for the valid problems,  $t(104)= -5.478$ ,  $p<.001$ , and for the invalid problems,  $t(104)= -2.627$ ,  $p<.01$ , were both significant, with the neutral conclusions being judged as valid more often both for invalid problems (*Mean difference* =.07, *SD* =.26) and for valid problems (*Mean difference* =.23, *SD* =.43).

### *Confidence Ratings*

The omnibus analysis for the confidence ratings revealed a single main effect of Beliefs,  $F(2,208)=6.158$ ,  $MSE=.216.308$ ,  $p=.003$ ,  $partial \eta^2=.056$ . The main effect of Task,  $F(1,104)=.560$ ,  $MSE=586.506$ ,  $p=.456$ ,  $partial \eta^2=.005$ , and the main effect of Logic,  $F(1,104)=2.969$ ,  $MSE=134.960$ ,  $p=.088$ ,  $partial \eta^2=.028$ , were not significant. In contrast, all two-way interactions (Task x Beliefs:  $F(2,208)=6.095$ ,  $MSE=199.589$ ,  $p=.003(.004)$ ,  $partial \eta^2=.055$ ; Task x Logic:  $F(1,104)=8.962$ ,  $MSE=202.063$ ,  $p=.003$ ,  $partial \eta^2=.079$ ; Beliefs x Logic:  $F(2,208)=27.645$ ,  $MSE=113.298$ ,  $p<.001$ ,  $partial \eta^2=.210$ ) and the three-way interaction,  $F(2,208)=11.044$ ,  $MSE=97.031$ ,  $p<.001$ ,  $partial \eta^2=.096$ , were significant. Means and standard deviations are reported in **Table 6**. As with the acceptance ratings, the next natural step was to analyze the data separately by task.

*Deductive reasoning.* In this 3 x 2 (Beliefs x Logic) repeated-measures ANOVA, both main effects of Beliefs,  $F(2,208)=7.250$ ,  $MSE=240.801$ ,  $p=.001$ ,  $partial \eta^2=.065$ , and Logic,  $F(1,104)=15.416$ ,  $MSE=126.981$ ,  $p<.001$ ,  $partial \eta^2=.129$ , as well as the interaction,  $F(2,208)=3.062$ ,  $MSE=89.762$ ,  $p=.049$ ,  $partial \eta^2=.029$ , were significant. Because of this interaction, three 2 x 2 (Beliefs x Logic) repeated measures ANOVA, comparing two levels of Beliefs at a time, were also conducted, both to compare the current results with the results of

Table 6

Experiment 2: Confidence Ratings

| Logic          | Task      |       |        |       |
|----------------|-----------|-------|--------|-------|
|                | Deductive |       | Causal |       |
|                | M         | SD    | M      | SD    |
| Valid          |           |       |        |       |
| Believable     | 89.0      | 12.51 | 86.5   | 14.38 |
| Non-Believable | 85.3      | 15.27 | 81.7   | 16.81 |
| Neutral        | 82.3      | 18.58 | 84.2   | 16.84 |
| Invalid        |           |       |        |       |
| Believable     | 83.2      | 15.46 | 79.9   | 20.34 |
| Non-Believable | 84.1      | 17.67 | 92.3   | 11.96 |
| Neutral        | 78.8      | 17.55 | 84.1   | 19.60 |

Experiment 1 and to better understand the differences between the three levels of Beliefs. First, similar to Experiment 1, the believable level was compared to the non-believable level of Beliefs. The sole significant main effect was that of Logic,  $F(1,104)=11.783$ ,  $MSE=110.202$ ,  $p=.001$ , *partial*  $\eta^2=.102$ , with invalid problems leading to lower confidence ratings than valid problems. Unlike what was suggested by Shynkaruk and Thompson (2006), believable conclusions did not lead to higher confidence ratings than non-believable ones, as there was no effect of Beliefs,  $F(1,104)=1.311$ ,  $MSE=142.916$ ,  $p=.255$ , *partial*  $\eta^2=.012$ . The interaction between Beliefs and Logic was also significant,  $F(1,104)=7.234$ ,  $MSE=75.987$ ,  $p=.008$ , *partial*  $\eta^2=.065$ . Second, the next two analyses involved the neutral level of Beliefs. Participants' confidence ratings on these problems were significantly lower than both those for the believable problems,  $F(1,104)=11.969$ ,  $MSE=267.909$ ,  $p=.001$ , *partial*  $\eta^2=.103$  and those for the non-believable ones,  $F(1,104)=5.918$ ,  $MSE=311.580$ ,  $p=.017$ , *partial*  $\eta^2=.054$ . Main effects of Logic were also present in the analysis

with the believable,  $F(1,104)=18.379$ ,  $MSE=124.799$ ,  $p<.001$ , *partial*  $\eta^2=.150$ , and the non-believable problems,  $F(1,104)=5.497$ ,  $MSE=108.721$ ,  $p=.021$ , *partial*  $\eta^2=.050$ . Finally, no interaction was found (believable x neutral:  $F(1,104)=1.729$ ,  $MSE=77.656$ ,  $p=.191$ , *partial*  $\eta^2=.016$ ; non-believable x neutral:  $F(1,104)=1.216$ ,  $MSE=115.642$ ,  $p=.273$ , *partial*  $\eta^2=.012$ ).

Overall, this simply means that the neutral problems led to the lowest ratings of confidence in the deductive reasoning task, potentially emphasizing the importance of content in guiding decisions in this type of reasoning.

*Causal Reasoning.* With this task, the 3 x 2 (Beliefs x Logic) repeated measures ANOVA showed a significant main effect of Beliefs,  $F(2,208)=4.583$ ,  $MSE=175.095$ ,  $p=.011(.016)$ , *partial*  $\eta^2=.042$ , no effect of Logic,  $F(1,104)=1.209$ ,  $MSE=210.568$ ,  $p=.274$ , *partial*  $\eta^2=.011$ , and a significant interaction between Beliefs and Logic,  $F(2,208)=32.587$ ,  $MSE=120.568$ ,  $p<.001$ , *partial*  $\eta^2=.239$ . When comparing each level of Beliefs pairwise in a series of 2 x 2 (Beliefs x Logic) repeated measures ANOVA, the results of the believable versus non-believable comparison showed a main effect of Beliefs,  $F(1,104)=15.829$ ,  $MSE=93.517$ ,  $p<.001$ , *partial*  $\eta^2=.132$ , in addition a significant interaction,  $F(1,104)=40.242$ ,  $MSE=191.592$ ,  $p<.001$ , *partial*  $\eta^2=.279$ , but no main effect of Logic,  $F(1,104)=2.361$ ,  $MSE=169.914$ ,  $p=.127$ , *partial*  $\eta^2=.022$ . As previously found, the confidence ratings were lower when the logic of the problem was in conflict rather than congruent with the beliefs used in the problem. Those differences were significant both for valid problems,  $t(104)=3.036$ ,  $p=.003$ , and invalid ones,  $t(104)=-7.177$ ,  $p<.001$ .

The next two comparisons involved the neutral level of Beliefs. Here, the pattern of confidence ratings was different from the analogous analyses in the deductive task. First, Logic was the only significant main effect, both when neutral problems were compared to believable problems,  $F(1,104)=6.130$ ,  $MSE=193.050$ ,  $p=.015$ , *partial*  $\eta^2=.056$ , and when neutral problems

were compared to non-believable problems,  $F(1,104)=16.052$ ,  $MSE=177.688$ ,  $p<.001$ , *partial*  $\eta^2=.134$ ). In addition, Beliefs interacted with Logic both in the believable versus neutral comparison,  $F(1,104)=14.173$ ,  $MSE=78.598$ ,  $p<.001$ , *partial*  $\eta^2=.120$ , and in the non-believable versus neutral comparison  $F(1,104)=32.375$ ,  $MSE=91.513$ ,  $p<.001$ , *partial*  $\eta^2=.237$ ). In both cases, the interaction was driven by the significant differences in confidence ratings at the invalid level of Logic. In fact, for the invalid condition, believable conclusions led to lower ratings of confidence than with the neutral problems,  $t(104)=-2.299$ ,  $p=.023$ . In contrast, non-believable conclusions led to higher confidence ratings than neutral problems at the invalid level of Logic,  $t(104)=5.147$ ,  $p<.001$ . No difference was found between the neutral level of beliefs and the belief-laden (both believable and non-believable items) conditions when the problems were valid. More importantly, the logic status of the problems did not seem to have any impact on the confidence ratings when the problems were neutral, as the mean ratings were comparable in both conditions (i.e., 84.2% versus 84.1%).

#### *Belief and Logic Across Reasoning Domains – Correlational Analyses*

To examine the use of beliefs and logic across tasks, difference scores were calculated the same way they were in Experiment 1. Specifically, adding the acceptance scores for the believable conditions and subtracting the sum of the scores in the non-believable conditions calculated belief indices for each participant. Similarly, the logic index was calculated by subtracting the sum of the scores on invalid problems from the sum of the scores on the valid problems. The correlations among these indexes are given in **Table 7**.

Unlike Experiment 1, the correlations between indexes yielded only one significant result. In the causal task, it was again found that a greater reliance on beliefs led to a lesser reliance on logic,  $r(103)=-.521$ ;  $p<.001$ . Unexpectedly, this relation was not found for the deductive task. No relation was found either between the uses of beliefs across the two tasks.

Table 7

*Intercorrelations Between Indexes of Belief/Logic Use and Individual Differences.*

| Index                | 1 | 2    | 3    | 4       | 5               | 6      | 7               | 8       | 9              | 10             |
|----------------------|---|------|------|---------|-----------------|--------|-----------------|---------|----------------|----------------|
| 1. Belief-D          | – | .178 | .114 | -.073   | -.189<br>(.054) | -.165  | -.180<br>(.067) | .221*   | -.223*         | -.059          |
| 2. Logic-D           | – | –    | .013 | -.018   | .081            | .178   | .212*           | -.248*  | -.009          | .142           |
| 3. Belief-C          | – | –    | –    | -.521** | -.095           | -.052  | -.031           | .087    | .056           | .027           |
| 4. Logic-C           | – | –    | –    | –       | .166            | .065   | -.080           | -.073   | .079           | .135           |
| Thinking Disposition |   |      |      |         |                 |        |                 |         |                |                |
| 5. PNI               | – | –    | –    | –       | –               | .349** | .195*           | -.101   | .237*          | .046           |
| 6. NC                | – | –    | –    | –       | –               | –      | .214*           | -.325** | .191           | .248*          |
| 7. CRT               | – | –    | –    | –       | –               | –      | –               | -.127   | .177<br>(.071) | .180<br>(.067) |
| 8. Dogma             | – | –    | –    | –       | –               | –      | –               | –       | -.120          | -.331**        |
| Cognitive Ability    |   |      |      |         |                 |        |                 |         |                |                |
| 9. Digits            | – | –    | –    | –       | –               | –      | –               | –       | –              | .279**         |
| 10. 3-min            | – | –    | –    | –       | –               | –      | –               | –       | –              | –              |

\*\* . Correlation is significant at the .01 level (two-tailed).

\* . Correlation is significant at the .05 level (two-tailed).

Note. Belief-D= Index of Belief Use for the Deductive Task; Logic-D= Index of Logic Use for the Deductive Task; Belief-C= Index of Belief Use for the Causal Task; Logic-C = Index of Logic Use for the Causal Task; PNI= Preference for Numerical Information; NC= Need for Cognition; CRT= Cognitive Reflection Test; Dogma= Dogmatism; Digits= Backward Digits Test; 3-min= Baddeley’s 3-minute test.

*Individual Differences.* Surprisingly, the only significant correlations were related to the deductive task (see **Table 7**). First, the CRT showed a significant positive correlation with the use of logic,  $r(103)=.212, p=.03$ , and a marginally significant negative correlation,  $r(103)=-.180, p=.067$ , with the use of beliefs. Given that CRT is hypothesized to reflect the use of the analytical System 2, these results are in line with the predictions outlined. Second, the PNI was also correlated negatively, although marginally, with the use of beliefs in the deductive task,  $r(103)=-.189; p=.054$ . This relation was predicted, but for the causal task. The last thinking disposition to show significant correlations with the indexes was the Dogmatism scale. As predicted, a higher degree of dogmatism led to a greater use of beliefs,  $r(103)=.221, p=.023$ , and to a lesser use of logic,  $r(103)=-.248, p=.011$  in the deductive reasoning task. Finally, concerning the measures of Cognitive Ability, a greater working memory capacity (as indexed by the Backward Digit Span) led to a decrease in the use of beliefs,  $r(103)=-.223, p=.022$ . It is surprising that the same score did not correlate positively with the use of logic.

### *Summary*

As with Experiment 1, the belief bias effect was found in both domains of reasoning. Also, the scenarios containing neutral levels of belief were accepted to a similar rate as the believable scenarios. This was not the case with confidence, as neutral problems induced lower overall ratings in the deductive task at both levels of Logic. In contrast, their rating was unaffected by Logic on the causal problems. Finally, dogmatism, a thinking disposition, was the single most predictive measure, correlating both negatively with the use of logic and positively with the use of beliefs. Two other measures, the CRT and the Backward Digit Span, correlated either positively with the use of logic (CRT) or negatively with the use of beliefs (Backward Digit Span), but not both.

## General Discussion

The focus of the two experiments presented here was to examine the commonalities and distinctions in the processing of deductive and causal reasoning problems. As with previous studies, the belief bias effect was present, although reduced when neutral problems were included in the second experiment. The effect of beliefs on logic also followed previous findings, with the largest effects being found for the invalid scenarios in the deductive reasoning task and for the valid scenarios in the causal reasoning task. It is important to note that the belief bias effect was replicated, albeit somewhat reduced, despite the use of a binary scale for the acceptance ratings in both tasks. Confidence ratings were also influenced by the combination of beliefs and logic, with conflict conditions leading to lower ratings than congruent conditions.

In the search of predicting factors, indexes of belief use and of logic use were calculated for each task. Whereas the two types of indexes correlated negatively within each task, neither belief use nor logic use was associated across the two reasoning tasks, with the exception of a moderate relationship in belief use between the two domains in Experiment 1. Six measures of individual differences were also analyzed in light of these indexes. Surprisingly, only dogmatism was predictive of both belief use and logic use in deductive reasoning. Also in relation to the deductive task, CRT correlated positively with the use of logic and the Backward Digit Span score correlated negatively with the use of beliefs. No measure of individual differences was predictive of the performance on the causal reasoning task.

The above findings will now be discussed in term of (1) reasoning ability and dual system theories, (2) the influence of introducing a neutral condition to the effects of Beliefs, and (3) the influence of beliefs/logic conflicts on confidence. Finally, I will consider the use of beliefs in research and some of the difficulties that arise from their utilization.

*Reasoning ability.* As with reaching the stage of formal operations (Inhelder & Piaget, 1958), deductive reasoning has long been seen as the pinnacle of reasoning, the ultimate intellectual achievement. Following this view, one would expect that being skillful at solving syllogisms would be predictive of the ability to reason rationally, based on logic and evidence, in multiple reasoning domains. This is not what was found in the two experiments. In fact, indexes of logic use on the deductive task and on the causal task did not correlate in either experiment. Inconsistent with what might be expected based on contemporary dual process theories and some scales, such as the CRT (Frederick, 2005), that may be designed to tap into individual tendencies to use either system preferentially, the same individuals did not display a reliable relationship between the use of beliefs or the use of logic across tasks. Neither a prominent reliance on their beliefs (i.e., System 1), or a consistent use of logic (i.e., System 2), was found to appear across both reasoning domains. Although it first appeared that the use of System 1 might have been more general, the correlation between the use of beliefs across tasks disappeared in the second experiment. In regard to the use of logic, the lack of correlation between the indexes from the two tasks seems to clearly indicate that the reliance on System 2 might be domain specific rather than domain general. This specificity in reasoning is line with previous findings (e.g., Bonnefon, Vautier & Eid, 2006).

Given the lack of any reliable relationship between the two domains, it might be useful to investigate the role that the limited judgment scale had on this result. If the concern remains to equate both tasks, one may choose to use a four-point scale across both tasks. For instance, Evans, Handley, Harper and Johnson-Laird (1999) did argue that the use of four-point scale (i.e., necessarily valid – possibly valid – possibly invalid – necessarily invalid) might reflect a more natural way in which participants evaluate and understand syllogistic problems. Of course, this change in scale would bring the causal task closer to its probabilistic nature.

*Finding an appropriate belief baseline: “Neutral” may not be so neutral.* The introduction of a neutral level of beliefs in Experiment 2 was done to better understand the role of beliefs, especially in their interaction with logic, in influencing acceptance ratings to create the belief bias effect. The main prediction that neutral and believable problems would be evaluated in a similar fashion was supported. On the other hand, the addition of a neutral level of beliefs in Experiment 2 appears to have had an influence on the processing of the belief-based problems. Indeed, the results obtained for believable and unbelievable problems were not perfectly replicated from Experiment 1. A potential explanation is the inclusion of the neutral problems with the other types of problems, as all problems were randomized to avoid any order effect, may have rendered the underlying structure of the other problems more salient, in turn influencing the response pattern of participants. Specifically, responding to “Neutral” scenarios, where one can only rely on the underlying logical structure, may impose a logical reasoning task set whereby content is ignored. To the extent that this may have occurred in the present experiment, it may have been preferable to block the neutral problems together and present them after the belief-laden problems. In the case where two different reasoning tasks are used in a within-subject-design, as in the present experiments, it would also be necessary to decide whether the neutral problems should be presented at the end of each section (e.g., causal-neutral problems after the causal problems) or after each task has been completed (i.e., both causal- and deductive-neutral problems after all other problems have been answered).

*Conflicts and confidence.* Unlike Shynkaruk and Thompson (2006), we focused our analysis not only on the influence of the believability of the conclusion on self-reported confidence, but on the influence of the interaction between beliefs and logic on confidence ratings. The first experiment clearly demonstrated that participants’ confidence in their answer dictated not only by the belief status of the conclusion, but rather by the combination between

beliefs and logic. In fact, participants showed sensitivity to conflicts between beliefs and logic, their ratings being lower in cases of conflict in comparison to non-conflict situations. This awareness should be emphasized as a metacognitive strategy to recognize situations in which one is more likely to be swayed by their prior beliefs and/or more easily discounting available evidence when judging the conclusion of an argument.

*Defining belief.* When studying the influence of prior beliefs on reasoning, many assumptions are made in the classification of stimuli as being either believable or non-believable. Common knowledge and intuition are often used as the basis for this classification, thereby making the generalization of findings limited. In fact, one difficulty in studying beliefs is their variability across the population (Rokeach, 1968) and their varying influence across domains (Toplak & Stanovich, 2003). Another difficulty relates to the variability of the strength of beliefs across individuals (Klaczynski, 2006). One strategy, used by Goel et Dolan (2001) as well as Stanovich and West (1997), is to present only the conclusion from each presented arguments to the participants and ask them to qualify each statement as true or false. As much as this is an appropriate strategy to find out the prior beliefs of participants, the use of a true/false scale, as was employed in the present experiments, will likely not represent the potential continuum of opinions on different topics. On the other hand, if, as suggested by MacPherson and Stanovich (2007), the belief bias applies only to common knowledge, whereas the preoccupation with opinions belongs to the realm of the Myside bias (e.g., Toplak & Stanovich, 2003), it may be more appropriate to keep working exclusively with knowledge when using the belief bias paradigm. Yet, given we often seek knowledge and information consistent with our opinions (e.g., confirmation bias), this differentiation between the two types of biases and their respective paradigms may not make complete sense.

*Conclusion.* Reasoning rarely occurs in a vacuum. We bring into the evaluation of any argument a certain number of assumptions as to what is true, what should be true, and what we want to be true. Relying on what we know and believe in evaluating problems and situations allows one to reach a solution or conclusion efficiently. It is easy, fast and often adaptive. Studying the belief bias effect highlights some of the situations in which letting our beliefs guide our decisions may lead to sub-optimal decisions. The risk of bias increases as the underlying assumptions of a problem contradict what we believe. If, as the current experiments demonstrated, people can be sensitive to their decrease in confidence as a sign of a potentially biased decision, then it might become a useful tool. This might be especially true in situations with high stakes, such as in the justice system or in any other process of evaluation (e.g., job interview, grading of assignments, contractual bids.). Finally, the requirement for many of those high stakes situations to employ personnel able to display some objectivity in their judgments could justify the use of measures of individual differences that would consistently be found to predict the use of prior beliefs to the detriment of using all incoming evidence to reach an informed decision.

# Appendix A

## Deductive Problems

### Instructions:

#### Logical Deductive Reasoning

This portion of the experiment is designed to find out how people solve logical problems. In the booklet that you have been given, there are twelve logical reasoning problems. Your task is to decide whether the conclusion below each problem follows logically from the information in the problem. **You must assume that all the information that you are given is true.** If, and only if, you judge that a given conclusion logically follows from the information given, you should circle "YES" beside that conclusion. If you think that the given conclusion does not necessarily follow from the given information, you should circle "NO". After each of your answers, please indicate on scale of 0 to 100 your degree of confidence in your answer (0-Not confident at all; 100-Absolutely certain).

**Please answer the problems in the order in which they appear, and ask any questions you have before you begin.**

-----



|  |  |
|--|--|
| On a scale of 0 to 100, how confident are you in your answer?  | On a scale of 0 to 100, how confident are you in your answer?  |
| <u>Cigarettes</u><br><b>Given that:</b><br>No addictive things are inexpensive<br>Some cigarettes are inexpensive<br><br><b>Does it necessarily follow that:</b><br>Some addictive things are not cigarettes<br><p style="text-align: center;">YES    NO</p> | <u>Fish</u><br><b>Given that:</b><br>No fish are colourful things<br>Some trout are colourful things<br><br><b>Does it necessarily follow that:</b><br>Some fish are not trout<br><p style="text-align: center;">YES    NO</p> |
| On a scale of 0 to 100, how confident are you in your answer?  | On a scale of 0 to 100, how confident are you in your answer?  |

**Deductive – Non-Believable – Valid**

| Version 1   | Version 2  |
|---|--|
| <u>Priests</u><br><b>Given that:</b><br>No religious people are married<br>Some priests are married<br><br><b>Does it necessarily follow that:</b><br>Some priests are not religious people<br><p style="text-align: center;">YES    NO</p> | <u>Drinks</u><br><b>Given that:</b><br>No drinks are sweet things<br>Some liquors are sweet things<br><br><b>Does it necessarily follow that:</b><br>Some liquors are not drinks<br><p style="text-align: center;">YES    NO</p> |
| On a scale of 0 to 100, how confident are you in your answer?   | On a scale of 0 to 100, how confident are you in your answer?  |
| <u>Buildings</u><br><b>Given that:</b><br>No buildings are wooden things<br>Some skyscrapers are wooden things<br><br><b>Does it necessarily follow that:</b>   | <u>Millionaires</u><br><b>Given that:</b><br>No millionaires are scientists<br>Some rich people are scientists<br><br><b>Does it necessarily follow that:</b>  |

|   |  |
|---|--|
| <p>Some skyscrapers are not buildings</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p>Some millionaires are not rich people</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |
|---|--|

**Deductive – Non-Believable – Invalid**

| Version 1   | Version 2  |
|---|--|
| <p><u>Drinks</u></p> <p><b>Given that:</b><br/>           No liquors are sweet things<br/>           Some drinks are sweet things</p> <p><b>Does it necessarily follow that:</b><br/>           Some liquors are not drinks</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>                       | <p><u>Priests</u></p> <p><b>Given that:</b><br/>           No priests are married<br/>           Some religious people are married</p> <p><b>Does it necessarily follow that:</b><br/>           Some priests are not religious people</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>         |
| <p><u>Millionaires</u></p> <p><b>Given that:</b><br/>           No millionaires are scientists<br/>           Some rich people are scientists</p> <p><b>Does it necessarily follow that:</b><br/>           Some millionaires are not rich people</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Buildings</u></p> <p><b>Given that:</b><br/>           No skyscrapers are wooden things<br/>           Some buildings are wooden things</p> <p><b>Does it necessarily follow that:</b><br/>           Some skyscrapers are not buildings</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |

**Deductive – Neutral – Valid**

| Version 1   | Version 2   |
|---|---|
| <p><u>T</u></p> <p><b>Given that:</b><br/>                     No T are D<br/>                     Some L are D</p> <p><b>Does it necessarily follow that:</b><br/>                     Some L are not T</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>F</u></p> <p><b>Given that:</b><br/>                     No F are O<br/>                     Some B are O</p> <p><b>Does it necessarily follow that:</b><br/>                     Some B are not F</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |
| <p><u>X</u></p> <p><b>Given that:</b><br/>                     No X are Z<br/>                     Some C are Z</p> <p><b>Does it necessarily follow that:</b><br/>                     Some C are not X</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>V</u></p> <p><b>Given that:</b><br/>                     No V are T<br/>                     Some X are T</p> <p><b>Does it necessarily follow that:</b><br/>                     Some X are not V</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |

**Deductive – Neutral– Invalid**

| Version 1  | Version 2  |
|--|--|
| <p><u>F</u></p> <p><b>Given that:</b><br/>                     No F are O<br/>                     Some B are O</p> <p><b>Does it necessarily follow that:</b></p> | <p><u>T</u></p> <p><b>Given that:</b><br/>                     No T are D<br/>                     Some L are D</p> <p><b>Does it necessarily follow that:</b></p> |

|  |  |
|--|--|
| <p>Some F are not B</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  | <p>Some T are not L</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  |
| <p><u>V</u></p> <p><b>Given that:</b><br/> No V are T<br/> Some X are T</p> <p><b>Does it necessarily follow that:</b><br/> Some V are not X</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>X</u></p> <p><b>Given that:</b><br/> No X are Z<br/> Some C are Z</p> <p><b>Does it necessarily follow that:</b><br/> Some X are not C</p> <p style="text-align: center;">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |

## Appendix B

### Causal Scenarios

#### **Instructions:**

#### **Causal Reasoning**

On the following pages you will be given a series of scenarios, each of which describes an event that has happened, and a possible cause for that event. Please read through the following scenarios carefully. **You must assume that all the information that you are given is true.** You must determine, based on the information given, if the given cause is responsible for the given effect. Record your answer on the yes/no scale. After each of your answers, please indicate on a scale of 0 to 100 your degree of confidence in your answer (0-Not confident at all; 100-Absolutely certain).

**Please respond to the scenarios in the order in which they occur. Once you have made a response to the scenario, please don't go back and change it. If you have any questions at this time, please feel free to ask the experimenter.**

-----

**Causal – Believable – Valid**

| Version 1   | Version 2  |
|---|--|
| <p><u>Depleted Fish</u></p> <p>Imagine you are a biologist who is trying to determine the cause of a recent decrease in the population of fish in Canada’s lakes. You have a hypothesis that the decrease in fish may be due to the recent introduction of insecticides designed to halt the birth of mosquito larvae. In order to test this theory, you decide to investigate 10 lakes with insecticides and 10 lakes without insecticides. A thorough investigation revealed the following information: of the 10 lakes that contained insecticides, 10 had depleted fish populations; of the 10 lakes that did not contain insecticides, 0 had depleted fish populations.</p> <p>Given the above information, do you think that the introduction of insecticides caused the recent decrease in the population of fish in Canada’s lakes?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Car Accidents</u></p> <p>Imagine you are a police officer who is trying to determine the cause of a recent surge in accidents on the Trans-Canada Highway. You have a hypothesis that the accidents may be due to severed brake lines. To test this theory, you examine 10 cars that had severed brake lines and 10 cars that did not have severed brake lines. A thorough investigation revealed the following information: of the 10 cars that had severed brake lines, 10 were involved in an accident; of the 10 cars that did not have severed brake lines, 0 were involved in an accident.</p> <p>Given the above information, do you think that severed brake lines caused the recent surge in accidents on the Trans-Canada Highway?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |
| <p><u>Fatigue</u></p> <p>Imagine you are a doctor trying to determine the cause of fatigue in a group of patients. You have a hypothesis that the fatigue may be due to insomnia. To test this theory, you investigated 10 patients who had insomnia and 10 patients who did not have insomnia. A thorough investigation revealed the following information: of the 10 patients that had insomnia, 10 were fatigued; of the 10 patients that did not have insomnia 0 were fatigued.</p> <p>Given the above information, do you think that insomnia caused the patients to be fatigued?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  | <p><u>Lung Cancer</u></p> <p>Imagine you are a scientist who is trying to determine the cause of the development of lung cancer in a group of patients. You have a hypothesis that the lung cancer may be due to smoking. To test this theory, you investigate 10 patients who smoked and 10 patients who did not smoke. A thorough investigation revealed the following information: of the 10 patients who smoked, 10 had lung cancer; of the 10 patients who did not smoke, 0 had lung cancer.</p> <p>Given the above information, do you think that smoking caused the development of lung cancer? YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>   |

**Causal – Believable – Invalid**

| Version 1  | Version 2   |
|--|---|
| <p><u>Car Accidents</u></p> <p>Imagine you are a police officer who is trying to determine the cause of a recent surge in accidents on the Trans-Canada Highway. You have a hypothesis that the accidents may be due to severed brake lines. To test this theory, you examine 10 cars that had severed brake lines and 10 cars that did not have severed brake lines. A thorough investigation revealed the following information: of the 10 cars that had severed brake lines, 0 were involved in an accident; of the 10 cars that did not have severed brake lines, 0 were involved in an accident.</p> <p>Given the above information, do you think that severed brake lines caused the recent surge in accidents on the Trans-Canada Highway?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Fatigue</u></p> <p>Imagine you are a doctor trying to determine the cause of fatigue in a group of patients. You have a hypothesis that the fatigue may be due to insomnia. To test this theory, you investigated 10 patients who had insomnia and 10 patients who did not have insomnia. A thorough investigation revealed the following information: of the 10 patients that had insomnia, 0 were fatigued; of the 10 patients that did not have insomnia 0 were fatigued.</p> <p>Given the above information, do you think that insomnia caused the patients to be fatigued?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  |
| <p><u>Lung Cancer</u></p> <p>Imagine you are a scientist who is trying to determine the cause of the development of lung cancer in a group of patients. You have a hypothesis that the lung cancer may be due to smoking. To test this theory, you investigate 10 patients who smoked and 10 patients who did not smoke. A thorough investigation revealed the following information: of the 10 patients who smoked, 0 had lung cancer; of the 10 patients who did not smoke, 0 had lung cancer.</p> <p>Given the above information, do you think that smoking caused the development of lung cancer? YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>   | <p><u>Depleted Fish</u></p> <p>Imagine you are a biologist who is trying to determine the cause of a recent decrease in the population of fish in Canada’s lakes. You have a hypothesis that the decrease in fish may be due to the recent introduction of insecticides designed to halt the birth of mosquito larvae. In order to test this theory, you decide to investigate 10 lakes with insecticides and 10 lakes without insecticides. A thorough investigation revealed the following information: of the 10 lakes that contained insecticides, 0 had depleted fish populations; of the 10 lakes that did not contain insecticides, 0 had depleted fish populations.</p> <p>Given the above information, do you think that the introduction of insecticides caused the recent decrease in the population of fish in Canada’s lakes?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |

**Causal – Non-Believable – Valid**

| Version 1   | Version 2  |
|---|--|
| <p><u>Flowers blooming</u><br/>Imagine you are a horticulturist who is trying to determine the cause of flowers blooming. You have a hypothesis that the flowers blooming may be due to playing classical music. To test this theory, you investigate 10 flowers that were surrounded by classical music and 10 flowers that were not surrounded by classical music. A thorough investigation revealed the following information: of the 10 flowers that were surrounded by classical music, 10 bloomed; of the 10 flowers that were not surrounded by classical music, 0 bloomed.</p> <p>Given the above information, do you think that playing classical music caused the flowers to bloom?<br/>YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Exam Success</u><br/>Imagine you are a professor who is trying to determine the cause of exam success. You have a hypothesis that exam success may be due to eating cornflakes for breakfast. To test this theory, you investigate 10 students who ate cornflakes for breakfast and 10 students who did not eat cornflakes for breakfast. A thorough investigation revealed the following information: of the 10 students who ate cornflakes for breakfast, 10 wrote a successful exam; of the 10 students who did not eat cornflakes for breakfast, 0 wrote a successful exam.</p> <p>Given the above information, do you think that eating cornflakes for breakfast caused exam success?<br/>YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>   |
| <p><u>Car Start Failures</u><br/>Imagine you are a tow-truck driver who is trying to determine the cause of a recent surge in car start failures. You have a hypothesis that the car start failures may be due to flat tires. To test this theory, you examine 10 cars that had flat tires and 10 cars that did not have flat tires. A thorough investigation revealed the following information: of the 10 cars that had flat tires, 10 failed to start; of the 10 cars that did not have flat tires, 0 failed start.</p> <p>Given the above information, do you think that flat tires caused the recent surge in car start failures?<br/>YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  | <p><u>Allergic Reaction</u><br/>Imagine you are a doctor who is trying to determine the cause of a recent surge of allergic reactions in children. You have a hypothesis that the allergic reactions may be due to doing homework. To test this theory, you examine 10 children who were doing homework prior to admission and 10 children who were not doing homework prior to admission. A thorough investigation revealed the following information: of the 10 children who were doing homework prior to admission, 10 were displaying signs of an allergic reaction; of the 10 children who were not doing homework prior to admission, 0 were displaying signs of an allergic reaction.</p> <p>Given the above information, do you think that doing homework caused the recent surge in allergic reactions in children?<br/>YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |

**Causal – Non-Believable – Invalid**

| Version 1   | Version 2  |
|---|--|
| <p><u>Exam Success</u><br/>Imagine you are a professor who is trying to determine the cause of exam success. You have a hypothesis that exam success may be due to eating cornflakes for breakfast. To test this theory, you investigate 10 students who ate cornflakes for breakfast and 10 students who did not eat cornflakes for breakfast. A thorough investigation revealed the following information: of the 10 students who ate cornflakes for breakfast, 0 wrote a successful exam; of the 10 students who did not eat cornflakes for breakfast, 0 wrote a successful exam.</p> <p>Given the above information, do you think that eating cornflakes for breakfast caused exam success?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>   | <p><u>Flowers blooming</u><br/>Imagine you are a horticulturist who is trying to determine the cause of flowers blooming. You have a hypothesis that the flowers blooming may be due to playing classical music. To test this theory, you investigate 10 flowers that were surrounded by classical music and 10 flowers that were not surrounded by classical music. A thorough investigation revealed the following information: of the 10 flowers that were surrounded by classical music, 0 bloomed; of the 10 flowers that were not surrounded by classical music, 0 bloomed.</p> <p>Given the above information, do you think that playing classical music caused the flowers to bloom?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |
| <p><u>Allergic Reaction</u><br/>Imagine you are a doctor who is trying to determine the cause of a recent surge of allergic reactions in children. You have a hypothesis that the allergic reactions may be due to doing homework. To test this theory, you examine 10 children who were doing homework prior to admission and 10 children who were not doing homework prior to admission. A thorough investigation revealed the following information: of the 10 children who were doing homework prior to admission, 0 were displaying signs of an allergic reaction; of the 10 children who were not doing homework prior to admission, 0 were displaying signs of an allergic reaction.</p> <p>Given the above information, do you think that doing homework caused the recent surge of allergic reactions in children?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Car Start Failures</u><br/>Imagine you are a tow-truck driver who is trying to determine the cause of a recent surge in car start failures. You have a hypothesis that the car start failures may be due to flat tires. To test this theory, you examine 10 cars that had flat tires and 10 cars that did not have flat tires. A thorough investigation revealed the following information: of the 10 cars that had flat tires, 0 failed to start; of the 10 cars that did not have flat tires, 0 failed start.</p> <p>Given the above information, do you think that flat tires caused the recent surge in car start failures?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  |

**Causal – Neutral – Valid**

| Version 1  | Version 2   |
|--|---|
| <p><u>Accountant</u></p> <p>Imagine you are an accountant who is trying to determine the cause of the <i>effect E</i> in the <i>population B</i>. You have a hypothesis that the <i>effect E</i> may be due to the <i>cause T</i>. In order to test this theory, you decide to investigate 10 <i>groups B</i> with the <i>cause T</i> and 10 <i>groups B</i> without the <i>cause T</i>. A thorough investigation revealed the following information: of the 10 <i>groups B</i> with the <i>cause T</i>, 10 had the <i>effect E</i>; of the 10 <i>groups B</i> that did not have the <i>cause T</i>, 0 had <i>effect E</i>.</p> <p>Given the above information, do you think that the <i>cause T</i> caused the <i>effect E</i> in <i>population B</i>?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Veterinarian</u></p> <p>Imagine you are a veterinarian who is trying to determine the cause of the <i>effect Y</i> in the <i>population F</i>. You have a hypothesis that the <i>effect Y</i> may be due to the <i>cause H</i>. In order to test this theory, you decide to investigate 10 <i>groups F</i> with the <i>cause H</i> and 10 <i>groups F</i> without the <i>cause H</i>. A thorough investigation revealed the following information: of the 10 <i>groups F</i> with the <i>cause H</i>, 10 had the <i>effect Y</i>; of the 10 <i>groups F</i> that did not have the <i>cause H</i>, 0 had <i>effect Y</i>.</p> <p>Given the above information, do you think that the <i>cause H</i> caused the <i>effect Y</i> in <i>population F</i>?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |
| <p><u>Engineer</u></p> <p>Imagine you are an engineer who is trying to determine the cause of the <i>effect V</i> in the <i>population F</i>. You have a hypothesis that the <i>effect V</i> may be due to the <i>cause R</i>. In order to test this theory, you decide to investigate 10 <i>groups F</i> with the <i>cause R</i> and 10 <i>groups F</i> without the <i>cause R</i>. A thorough investigation revealed the following information: of the 10 <i>groups F</i> with the <i>cause R</i>, 10 had the <i>effect V</i>; of the 10 <i>groups F</i> that did not have the <i>cause R</i>, 0 had <i>effect V</i>.</p> <p>Given the above information, do you think that the <i>cause R</i> caused the <i>effect V</i> in <i>population F</i>?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>     | <p><u>Electrician</u></p> <p>Imagine you are an electrician who is trying to determine the cause of the <i>effect K</i> in the <i>population R</i>. You have a hypothesis that the <i>effect K</i> may be due to the <i>cause S</i>. In order to test this theory, you decide to investigate 10 <i>groups R</i> with the <i>cause S</i> and 10 <i>groups R</i> without the <i>cause S</i>. A thorough investigation revealed the following information: of the 10 <i>groups R</i> with the <i>cause S</i>, 10 had the <i>effect K</i>; of the 10 <i>groups R</i> that did not have the <i>cause S</i>, 0 had <i>effect K</i>.</p> <p>Given the above information, do you think that the <i>cause S</i> caused the <i>effect K</i> in <i>population R</i>?</p> <p align="center">YES    NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  |

**Causal – Neutral– Invalid**

| Version 1   | Version 2  |
|---|--|
| <p><u>Veterinarian</u></p> <p>Imagine you are a veterinarian who is trying to determine the cause of the <i>effect Y</i> in the <i>population F</i>. You have a hypothesis that the <i>effect Y</i> may be due to the <i>cause H</i>. In order to test this theory, you decide to investigate 10 <i>groups F</i> with the <i>cause H</i> and 10 <i>groups F</i> without the <i>cause H</i>. A thorough investigation revealed the following information: of the 10 <i>groups F</i> with the <i>cause H</i>, 0 had the <i>effect Y</i>; of the 10 <i>groups F</i> that did not have the <i>cause H</i>, 0 had <i>effect Y</i>.</p> <p>Given the above information, do you think that the <i>cause H</i> caused the <i>effect Y</i> in <i>population F</i>?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> | <p><u>Accountant</u></p> <p>Imagine you are an accountant who is trying to determine the cause of the <i>effect E</i> in the <i>population B</i>. You have a hypothesis that the <i>effect E</i> may be due to the <i>cause T</i>. In order to test this theory, you decide to investigate 10 <i>groups B</i> with the <i>cause T</i> and 10 <i>groups B</i> without the <i>cause T</i>. A thorough investigation revealed the following information: of the 10 <i>groups B</i> with the <i>cause T</i>, 0 had the <i>effect E</i>; of the 10 <i>groups B</i> that did not have the <i>cause T</i>, 0 had <i>effect E</i>.</p> <p>Given the above information, do you think that the <i>cause T</i> caused the <i>effect E</i> in <i>population B</i>?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p> |
| <p><u>Electrician</u></p> <p>Imagine you are an electrician who is trying to determine the cause of the <i>effect K</i> in the <i>population R</i>. You have a hypothesis that the <i>effect K</i> may be due to the <i>cause S</i>. In order to test this theory, you decide to investigate 10 <i>groups R</i> with the <i>cause S</i> and 10 <i>groups R</i> without the <i>cause S</i>. A thorough investigation revealed the following information: of the 10 <i>groups R</i> with the <i>cause S</i>, 0 had the <i>effect K</i>; of the 10 <i>groups R</i> that did not have the <i>cause S</i>, 0 had <i>effect K</i>.</p> <p>Given the above information, do you think that the <i>cause S</i> caused the <i>effect K</i> in <i>population R</i>?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>  | <p><u>Engineer</u></p> <p>Imagine you are an engineer who is trying to determine the cause of the <i>effect V</i> in the <i>population F</i>. You have a hypothesis that the <i>effect V</i> may be due to the <i>cause R</i>. In order to test this theory, you decide to investigate 10 <i>groups F</i> with the <i>cause R</i> and 10 <i>groups F</i> without the <i>cause R</i>. A thorough investigation revealed the following information: of the 10 <i>groups F</i> with the <i>cause R</i>, 0 had the <i>effect V</i>; of the 10 <i>groups F</i> that did not have the <i>cause R</i>, 0 had <i>effect V</i>.</p> <p>Given the above information, do you think that the <i>cause R</i> caused the <i>effect V</i> in <i>population F</i>?</p> <p align="center">YES NO</p> <p>On a scale of 0 to 100, how confident are you in your answer?</p>     |

## Appendix C

### Thinking Dispositions

**INSTRUCTIONS:** You will read a list of statements. For each statement, use the numbers on the scale below to indicate your level of agreement or disagreement with that statement.

| <b>1</b>                 | <b>2</b>        | <b>3</b>                 | <b>4</b>                          | <b>5</b>              | <b>6</b>     | <b>7</b>              |
|--------------------------|-----------------|--------------------------|-----------------------------------|-----------------------|--------------|-----------------------|
| <i>Strongly disagree</i> | <i>Disagree</i> | <i>Slightly disagree</i> | <i>Neither agree nor disagree</i> | <i>Slightly agree</i> | <i>Agree</i> | <i>Strongly agree</i> |

#### Need for Cognition Scale

1. I would prefer complex problems to simple ones. \_\_\_\_\_
2. I like to have the responsibility of handling a situation that requires a lot of thinking.  
\_\_\_\_\_
3. Thinking is not my idea of fun. \_\_\_\_\_
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. \_\_\_\_\_
5. I try to anticipate and avoid situations where there is likely chance I will have to think in depth about something. \_\_\_\_\_
6. I find satisfaction in deliberating hard and for long hours. \_\_\_\_\_
7. I only think as hard as I have to. \_\_\_\_\_
8. I prefer to think about small, daily projects to long-term ones. \_\_\_\_\_
9. I like tasks that require little thought once I've learned them. \_\_\_\_\_
10. The idea of relying on thought to make my way to the top appeals to me. \_\_\_\_\_
11. I really enjoy a task that involves coming up with new solutions to problems. \_\_\_\_\_
12. Learning new ways to think doesn't appeal to me very much. \_\_\_\_\_
13. I prefer my life to be filled with puzzles that I must solve. \_\_\_\_\_
14. The notion of thinking abstractly is appealing to me. \_\_\_\_\_
15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought. \_\_\_\_\_

16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort. \_\_\_\_\_
17. It's enough for me that something gets the job done; I don't care how or why it works. \_\_\_\_\_
18. I usually end up deliberating about issues even when they do not affect me personally. \_\_\_\_\_

### **Preference for Numerical Information Scale**

1. I enjoy work that requires the use of numbers. \_\_\_\_\_
2. I think quantitative information is difficult to understand. \_\_\_\_\_
3. I find it satisfying to solve day-to-day problems involving numbers. \_\_\_\_\_
4. Numerical information is very useful in everyday life. \_\_\_\_\_
5. I prefer not to pay attention to information involving numbers. \_\_\_\_\_
6. I think more information should be available in numerical form. \_\_\_\_\_
7. I don't like to think about issues involving numbers. \_\_\_\_\_
8. Numbers are not necessary for most situations. \_\_\_\_\_
9. Thinking is enjoyable when it does not involve quantitative information. \_\_\_\_\_
10. I like to make calculations using numerical information. \_\_\_\_\_
11. Quantitative information is vital for accurate decisions. \_\_\_\_\_
12. I enjoy thinking about issues that do not involve numerical information. \_\_\_\_\_
13. Understanding numbers is as important in daily life as reading or writing. \_\_\_\_\_
14. I easily lose interest in graphs, percentages, and other quantitative information. \_\_\_\_\_
15. I don't find numerical information to be relevant for most situations. \_\_\_\_\_
16. I think it is important to learn and use numerical information to make well-informed decisions. \_\_\_\_\_
17. Numbers are redundant for most situations. \_\_\_\_\_
18. It is a waste of time to learn information containing a lot of numbers. \_\_\_\_\_
19. I like to go over numbers in my mind. \_\_\_\_\_
20. It helps me to think if I put down information as numbers. \_\_\_\_\_

### Dogmatism Scale

1. The United States and Russia have just about nothing in common. \_\_\_\_\_
2. The highest form of government is a democracy and the highest form of democracy is a government run by those who are most intelligent. \_\_\_\_\_
3. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups. \_\_\_\_\_
4. Man on his own is a helpless and miserable creature. \_\_\_\_\_
5. Most people just don't give a "damn" for others. \_\_\_\_\_
6. I'd like it if I could find someone who would tell me how to solve my personal problems. \_\_\_\_\_
7. In a discussion I often find it necessary to repeat myself several times to make sure I am being understood. \_\_\_\_\_
8. It is better to be a dead hero than to be a live coward. \_\_\_\_\_
9. While I don't like to admit this even to myself, my secret ambition is to become a great man, like Einstein, or Beethoven, or Shakespeare. \_\_\_\_\_
10. The main thing in life is for a person to want to do something important. \_\_\_\_\_
11. It is only when a person devotes herself to an ideal or cause that life becomes meaningful. \_\_\_\_\_
12. Of all the different philosophies that exist in this world there is probably only one that is correct. \_\_\_\_\_
13. To compromise with our political opponents is dangerous because it usually leads to the betrayal of our own side. \_\_\_\_\_
14. There are two kinds of people in this world: those who are for the truth and those who are against the truth. \_\_\_\_\_
15. My blood boils whenever a person stubbornly refuses to admit she's wrong. \_\_\_\_\_
16. Most of the ideas that get printed nowadays aren't worth the paper they are printed on. \_\_\_\_\_
17. In this complicated world of ours the only way we can know what's going on is to rely on leaders or experts who can be trusted. \_\_\_\_\_
18. It is often desirable to reserve judgment about what's going on until one has had a chance to hear the opinions of those one respects. \_\_\_\_\_
19. The present is all too often full of unhappiness. It is only the future that counts. \_\_\_\_\_
20. Most people just don't know what's good for them. \_\_\_\_\_

## Cognitive Reflection Test (CRT)

### Instructions:

You will be presented with three short problems. Answer them in order. There is space on each page to write down your answer.

### Questions:

(1) 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? \_\_\_\_\_ cents

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? \_\_\_\_\_ minutes

(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 the lake?  
\_\_\_\_\_ days

## Appendix D

### Cognitive Ability

#### Baddeley's 3-Minute Test

##### Instructions:

In the following test there are a number of short sentences each followed by a pair of letters (AB or BA). The sentences claim to describe the order of the two letters, i.e. to say which comes first. They can do so in several different ways. Thus the order AB can be correctly described by saying either (1) A precedes B, or (2) B follows A, or (3) B does not precede A, or (4) A does not follow B. All these are correct descriptions of the pair AB but are incorrect when applied to the other pair BA.

Your task is to read each sentence and to decide whether it is a true or a false description of the letter pair that follows it. If you think that the sentence describes the letter pair correctly put a tick in the first column (labelled "True"). If you think the sentence does not give a correct description of the letter order, put a tick in the second ("False") column.

Here is an example:

|                      | TRUE                                | FALSE                               |
|----------------------|-------------------------------------|-------------------------------------|
| 1. A precedes B – AB | <input checked="" type="checkbox"/> |                                     |
| 2. B precedes A – AB |                                     | <input checked="" type="checkbox"/> |

When you start the main test, work as quickly as you can without making mistakes. **Start with sentence one and work systematically through the test leaving no blank spaces.** The experimenter will let you know when to start and when to stop. Please ask any question that you may have at this time.

## Participant Questionnaire

|                                 | TRUE | FALSE |
|---------------------------------|------|-------|
| 1. A precedes B – AB            |      |       |
| 2. B is not followed by A – BA  |      |       |
| 3. B is preceded by A – AB      |      |       |
| 4. B is followed by A – AB      |      |       |
| 5. A is followed by B – AB      |      |       |
| 6. B does not follow A – BA     |      |       |
| 7. A is not followed by B – AB  |      |       |
| 8. A does not follow B – AB     |      |       |
| 9. B does not precede A – AB    |      |       |
| 10. B follows A – AB            |      |       |
| 11. A follows B – BA            |      |       |
| 12. A is followed by B – BA     |      |       |
| 13. A is preceded by B – BA     |      |       |
| 14. A is preceded by B – AB     |      |       |
| 15. B is followed by A – BA     |      |       |
| 16. A is not preceded by B – BA |      |       |
| 17. B is not followed by A – AB |      |       |
| 18. A does not precede B – AB   |      |       |
| 19. B precedes A – BA           |      |       |
| 20. B precedes A – AB           |      |       |
| 21. B does not precede A – BA   |      |       |
| 22. A precedes B – BA           |      |       |
| 23. B is not preceded by A – BA |      |       |
| 24. B is preceded by A – BA     |      |       |
| 25. A does not precede B – BA   |      |       |
| 26. A is not followed by B – BA |      |       |
| 27. B is not preceded by A – AB |      |       |
| 28. B follows A – BA            |      |       |
| 29. B does not follow A – AB    |      |       |
| 30. A is not preceded by B – AB |      |       |
| 31. A does not follow B – BA    |      |       |
| 32. A follows B – AB            |      |       |

## Backward Digit Span

### Researcher Instructions:

Background Description: The experimenter presents digits starting with sequence length 2 and increasing to a maximum of 9 orally. The participant must repeat the digits immediately, in reverse order. The task is discontinued when the participant makes two consecutive failures. The highest number of correctly recalled digits defines the backward digit span score (i.e. between 2 and 9).

### Instructions (to be read to the participant):

“I will read out loud a sequence of digits. Your task is to tell me the digits in reverse order. For example, if I say ‘4 – 3 – 7’, the correct answer would be to say ‘7 – 3 – 4’. What would your answer be if I say ‘1 – 3’? (*Let the participant say ‘3 – 1’*). That is correct. Do you have any questions about this task? (*Let the participant enough time to give you a clear No... Otherwise, clarify any question before beginning.*) Then, let’s begin.”

### Scoring:

- 1) Read the sequence slowly and clearly.
- 2) On the answer sheet, write down the participant’s answer (each digit s/he mentions).
- 3) Compare the participant’s answer to the answer key (printed on the answer sheet).
- 4) Indicate whether or not the sequence was correctly repeated in reverse order (correct: check under ‘Pass’; incorrect: check under ‘Fail’).
- 5) The task is discontinued when the participant makes two consecutive failures.
- 6) The participant’s score is defined by the highest number of correctly recalled digits (i.e. between 2 and 9).

**Digits To Be Read:**

2) **4-9**

2) **6-8**

3) **5-6-4**

3) **3-7-1**

4) **5-9-6-1**

4) **8-2-9-6**

5) **7-4-1-9-2**

5) **2-6-1-5-7**

6) **1-9-4-8-6-7**

6) **8-6-9-1-7-2**

7) **5-1-4-2-9-6-8**

7) **6-2-5-9-3-8-7**

8) **1-4-3-9-2-5-6-8**

8) **8-3-1-9-6-4-5-2**

9) **1-4-7-2-8-3-6-5-9**

9) **4-5-3-6-1-7-2-9-8**

**Backward Digit Span Response Sheet:**

Participant #: \_\_\_\_\_ Date: \_\_\_\_\_

RA: \_\_\_\_\_

(\* Stop testing after two "fail" in a row)

| Correct answers        | Participant's answers | Pass | Fail |
|------------------------|-----------------------|------|------|
| 2) [9-4]               |                       |      |      |
| 2) [8-6]               |                       |      |      |
|                        |                       |      |      |
| 3) [4-6-5]             |                       |      |      |
| 3) [1-7-3]             |                       |      |      |
|                        |                       |      |      |
| 4) [1-6-9-5]           |                       |      |      |
| 4) [6-9-2-8]           |                       |      |      |
|                        |                       |      |      |
| 5) [2-9-1-4-7]         |                       |      |      |
| 5) [7-5-1-6-2]         |                       |      |      |
|                        |                       |      |      |
| 6) [7-6-8-4-9-1]       |                       |      |      |
| 6) [2-7-1-9-6-8]       |                       |      |      |
|                        |                       |      |      |
| 7) [8-6-9-2-4-1-5]     |                       |      |      |
| 7) [7-8-3-9-5-2-6]     |                       |      |      |
|                        |                       |      |      |
| 8) [8-6-5-2-9-3-4-1]   |                       |      |      |
| 8) [2-5-4-6-9-1-3-8]   |                       |      |      |
|                        |                       |      |      |
| 9) [9-5-6-3-8-2-7-4-1] |                       |      |      |
| 9) [8-9-2-7-1-6-3-5-4] |                       |      |      |
|                        |                       |      |      |

Score (range: 2 to 9): \_\_\_\_\_

## References

- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*(3), 299-352.
- Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, *10*(10), 341-342.
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*(4), 553-573.
- Blais, A., Thompson, M. M., & Baranski, J. V. (2005). Individual differences in decision processing and confidence judgments in comparative judgment tasks: The role of cognitive styles. *Personality and Individual Differences*, *38*(7), 1701-1713.
- Bonnefon, J., Vautier, S., & Eid, M. (2007). Modeling individual differences in contrapositive reasoning with continuous latent state and trait variables. *Personality and Individual Differences*, *42*(7), 1403-1413.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55-81.
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*(4), 545-567.

- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90(2), 272-292.
- Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin*, 29(9), 1120-1132.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49(8), 709-724.
- Evans, J. S., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3), 295-306.
- Evans, J. S. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, 128(6), 978-996.
- Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454-459.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11(4), 382-389.

- Evans, J. S. B. T., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning*, 25(6), pp. 1495-1513, November 1999.
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Oxford, England: Psychology/Erlbaum (Uk) Taylor & Fr.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25(18)-43.
- Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, 43(8), 1204-1213.
- Fugelsang, J. A., & Thompson, V. A. (2000). Strategy selection in causal reasoning: When beliefs and covariation collide. *Canadian Journal of Experimental Psychology*, 54(1), 15-32.
- Fugelsang, J. A., & Thompson, V. A. (2001). Belief-based and covariation-based cues affect causal discounting. *Canadian Journal of Experimental Psychology*, 55(1), 70-76.
- Fugelsang, J. A., & Thompson, V. A. (2003). A dual-process model of belief and evidence interactions in causal reasoning. *Memory & Cognition*, 31(5), 800-815.
- Goel, V., & Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia*, 39(9), 901-909.

- Greenhoot, A. F., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology, 18*(2), 203-221.
- Hume, D. (Ed.). (1739/1978). *A treatise of human nature*. Oxford: Oxford University Press.
- Inhelder, B., Piaget, J., Parsons, A. (. , & Milgram, S. (. (1958). *The growth of logical thinking: From childhood to adolescence*. New York, NY, US: Basic Books.
- Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation, 15*, 192-238.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*(2), 107-128.
- Klaczynski, P. A. (2006). Learning, belief biases, and metacognition. *Journal of Cognition and Development, 7*(3), 295-300.
- 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.
- Klein, G. A. (1993). *A recognition-primed decision (RPD) model of rapid decision making*. Westport, CT, US: Ablex Publishing.

- Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development, 60*(6), 1316-1327.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*(4), 560-572.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior, 13*(5), 585-589.
- 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.
- Michotte, A. (1946). *La perception de la causalite. (etudes psychol. vol. VI.) / perception of causality*. Oxford, England: Inst. Sup. De Philosophie.
- Newstead, S. E., Pollard, P., Evans, J. S., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition, 45*(3), 257-284.
- Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farelly, D. (2004). Individual differences in deductive reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57*(1), 33-60.

- Rapus, T. L. (2004). Integrating information about mechanism and covariation in causal reasoning. ProQuest Information & Learning). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 65 (2-B) (Electronic; Print)
- Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 54(4), 1031-1048.
- Rokeach, M. (1960). *The open and closed mind*. Oxford, England: Basic Books.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, 10, 174–220.
- Sà, W. C., Kelley, C. N., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences*, 38(5), 1149-1161.
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299-310.
- Shaklee, H., & Elek, S. (1988). Cause and covariate: Development of two related concepts. *Cognitive Development*, 3(1), 1-13.

- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1), 1-51.
- Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34*(3), 619-632.
- Slovan, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.
- Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*(3), 165-173.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108-131.
- 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*(2), 342-357.
- Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188.

- Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, 38(3), 349-385.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5), 645-726.
- Thompson, V. A. (1996). Reasoning from false premises: The role of soundness in making logical deductions. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, 50(3), 315-319.
- Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, 17(7), 851-860.
- Troidahl, V. C., & Powell, F. A. (1965). A short-form dogmatism scale for use in field studies. *Social Forces*, 44(2), 211-214.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
- Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54(1), 68-80.

- Viswanathan, M. (1993). Measurement of individual differences in preference for numerical information. *Journal of Applied Psychology, 78*(5), 741-752.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology, 12*, 129-140.
- Wason, P. C., & Kosviner, A. (1966). Perceptual distortion induced by reasoning. *British Journal of Psychology, 57*(3-4), 413-418.
- Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior, 46*(1), 15-35.
- Wechsler, D. (Ed.). (1997). *Wechsler adult intelligence Scale—3rd edition (WAIS-3®)*. San Antonio, TX: Harcourt Assessment.
- White, P. A. (1988). Causal processing: Origins and development. *Psychological Bulletin, 104*(1), 36-52.
- White, P. A. (1989). A theory of causal processing. *The British Journal of Psychology, 80*, 431-454.
- Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology, 102*, 83.

Young, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin & Review*, 2(1), 83-104.