Logical Intuitions and Heuristic Reflections: Rethinking the Role of Intuition in Probability Judgements

Jennifer FAURE

Department of Management

Kingston University, London

March 2019

This thesis is submitted in partial fulfilment of the requirements of Kingston University

for the degree of Doctor of Philosophy

Table of contents

Abstract

This thesis aims to further our understanding of the role that intuition plays in human reasoning when making probability judgements. It attempts to: a) gain a better understanding of the cognitive processes underlying these judgements, b) determine how individual differences impacts the logicality of these judgements, and c) test original and theoretically-driven ways to increase logical intuitions in probability judgements. Classically, it is assumed that people make biased judgements because they rely on an intuitive thinking system (System 1) and apply the representativeness heuristic to make conjunctive probability judgements. In contrast, logical judgements are assumed to arise from the use of deliberation (System 2) to overrule the prepotent heuristic response and replace it with a logical one. Recent research; however, has challenged this claim and instead proposes that our intuitions do not always lead us astray. In fact, they can reflect a sensitivity to logic that is implicit, and potentially happens automatically and outside of awareness.

This thesis takes this notion one step further and asks whether it is the slower, more deliberative, thinking system which may be vulnerable to prior beliefs and biases. A series of five experiments examined the relative impact of heuristic and logical considerations on probability judgements. The results indicated that people are readily able to detect the conflict underlying intuitive and deliberative assessments, and that people effortlessly engage in deliberative processing, which suggests they are not simply cognitive misers who fail to reason in line with the principles of logic because they either lack the cognitive ability or the motivation to do so. The results also supported the idea that people can intuit logical judgements (i.e., judgements in accordance with the laws of probability) when they rely on System 1 thinking; however, when they deliberate or use System 2 thinking, that is when the heuristic biases their judgements.

List of Figures

List of Tables

Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Gaëlle Vallée-Tourangeau for sharing her immense expertise and knowledge with me. She has shown me by example how to be a good scientist. I feel privileged to have been mentored by such a skilled, enthusiastic, generous and awe-inspiring female role model. I would also like to extend my gratitude to Professor Frédéric Vallée-Tourangeau and Dr Sabira Mannan who provided invaluable insight and expertise into the design, implementation and interpretation of my research. Thank you to my colleague and friend Niyat Henok. You were always there when I needed you, whether as a shoulder to cry on or some light comic relief. Finally, I would like to thank my family who never stopped believing in me. Thank you to my husband who walked this journey alongside me, picking up all the pieces I dropped along the way. You have always been supportive of my dreams and have worked tirelessly to provide me with the time I needed to pursue them. Finally, thank you to my daughter for bringing immense joy and inspiration into my life. Spending time with you is always my favourite part of the day. This work was supported by a PhD studentship from the Leverhulme Trust [Research grant RPG-228].

Chapter 1: General Introduction

Judgement and decision making is a field that explores how people make decisions and judgements under uncertainty. It is a multidisciplinary area that bring together scientists and researchers from an array of disciplines including psychology, statistics, economics, philosophy, law and medicine to name just a few. Although the terms 'reasoning', 'judgement' and 'decision making' are usually used interchangeably, there are small differences that set them apart from each other. When people reason, they draw implications from their beliefs. This process resembles what philosophy describes as deduction: when someone seeks to derive new assertions from assertions already in place. For example, 'If I believe *X*, what other claims follow from this?' Reasoning allows us to use our knowledge in new ways. When people make judgements, they draw conclusions from their experiences. In other words, they draw new information from their prior experiences. Judgement and reasoning allow people to expand their knowledge in important ways. For example, when they draw new conclusions from their experiences, or when they deduce a novel claim from their other beliefs. When people make decisions, they choose from among options. This type of thinking is more closely linked to their actions and is informed by their judgements. The contents of thought and the processes of thought are linked in important ways, and the flow of their thoughts depends on what they are trying to accomplish in their thinking. The act of making a judgement involves reaching beyond the evidence gathered to draw a conclusion. Hence, the nature of making a judgement involves exploration and uncertainty because it requires reaching beyond what is already known; as well as risk because they might be mistaken in their judgement.

People's decision-making can be less than optimal at times and is prone to biases such as framing (i.e., the way a decision is phrased, or the way options are described can influence decisions by changing the point of reference). Judgements are also prone to biases and it is commonly assumed that these biases occur due to the use of heuristics. Two biases (i.e., the conjunction fallacy and base rate neglect) as well as the representativeness heuristic will be covered in more detail later in this chapter (Kahneman & Tversky, 1972; Tversky & Kahneman, 1974, 1982, 1983).

Probability

This thesis focusses on social judgments under uncertainty, specifically probability judgements, and probability is a concept that we use to deal with uncertainty. "The probability of an event is a measure of how likely it is to occur" (O'Hagan et al., 2006, p. 2). A probability of 0 means that the event will certainly not occur, while a probability of 1 means that it is certain to occur. Values increasing from 0 to 1 describe increasing chances that the event will occur, while the central value 0.5, represents an event that is as likely to occur as it is not to occur. For example, the toss of a coin. There is an equal chance of throwing a 'Heads' as there is to throw a 'Tails', thus if H is the probability of getting 'Heads', then the equation can be expressed as such: $P(\mathrm{H}) = 0.5$. Similarly, when rolling a die there are six equally likely results, thus if S is the probability of rolling a six, then $P(\mathrm{S}) = \frac{1}{6}$. The theoretical study of probability is a branch of mathematics that deals with laws and theorems about how probabilities behave and combine. However, when dealing with judgements under uncertainty it becomes a bit trickier as we cannot simply assign a numeric probability to an event such as we can with a coin toss or the roll of a die. Uncertainty is an unavoidable aspect of human existence and many important choices must be based on beliefs about the likelihood of such uncertain events occurring (Kahneman & Tversky, 1983). A few examples include: where to invest one's money, the result of an election, the outcome of a medical procedure, or the guilt of a defendant.

Rules of probability. There are many rules of probability all used to calculate the chance of an event happening. The rule of *mutual exclusivity* states that two events cannot both be true if they are mutually exclusive. For example, a) "Hillary Clinton will win the US presidency in 2016" and b) "Donald Trump will win the US presidency in 2016" are two statements each with a probability between 0 and 1; however, they both cannot be true because they are mutually exclusive. The rule of *exhaustiveness* is similar in that one of the probabilities must be true; however, when taken together they exhaust the possibilities. For example, a) "Einstein was born before 1900", b) "Einstein was born after 1900", and c) "Einstein was not born before or after 1900". Not only are these statements mutually exhaustive, one of them must be true; and together they exhaust the possibilities. The probability of *negations* enables one to calculate the probability of a negation from the probability of the statement negated. For example, suppose we know the probability of throwing a four on a die is 1 in 6, so $P(\mathrm{F}) = \frac{1}{6}$. Then the negation rule enables us to calculate the probability that a four will *not* turn up on the next throw: $P({\sim}\mathrm{F}) = \frac{5}{6}$. Obviously not every pair of statements is mutually exclusive, and in many cases two events can both be true. Thus, the general *disjunction* rule is used to calculate these probabilities. For example, the likelihood of drawing a King and drawing a Club from a deck of cards is not mutually exclusive because of the King of Clubs. So, one needs to subtract the probability of drawing that card: $P(\mathrm{K\&C}) = P(\mathrm{K}) + P(\mathrm{C}) - P(\mathrm{KC}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$. The *conditional* rule of probability enables one to calculate the probability that an event is true conditional on another event being true: $P(\mathrm{A/B}) = \frac{P(AB)}{P(B)}$. The rule of *independence* states that two events are independent if neither affect the probability of the other: $P(\mathrm{A/B}) = P(\mathrm{A})$ and $P(\mathrm{A/B}) = P(\mathrm{B})$.

For example, "Hillary Clinton will be the next US President" is independent of "the first card I choose (from a full deck) will be an Ace". When we are dealing with independent propositions, we can derive a simpler rule for conjunctions, the *restricted conjunction rule*: P(A&B) = P(A) x P(B). The final, but most applicable rule of probability for this thesis is the *conjunction probability rule*, as this rule forms the foundation on which this program of research was built.

Conjunction probability rule. Probability is a measurement of the likelihood that an event will occur. When making a judgement of probability, people are required to reach beyond what they already know, which adds an element of uncertainty and risk. The laws of probability are derived from extensional considerations. The extension rule is one of the simplest yet fundamental qualitative laws of probability. It states that if the extension of A includes the extension of B (i.e., A ⊃ B) then $P(\text{A}) \geq P(\text{B})$. This is because the set of possibilities associated with the conjunction A&B is included in the set of possibilities associated with B. This sentiment is also expressed in the conjunction rule $P(\text{A\&B}) \leq P(\text{A})$ and $P(\text{A\&B}) \leq P(B)$, which states that a conjunction can never be more likely to occur than one of its constituents (see Figure 1. 1). Even as the probability of B increases to cover A, it can still not be greater than A. At best, B can only be equal to A.

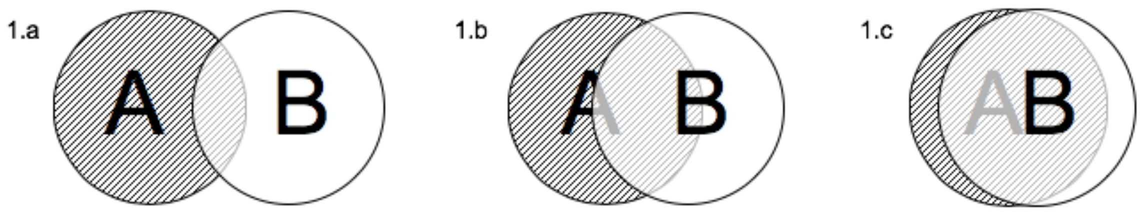


Figure 1. 1. The conjunction rule of probability as represented by Venn diagrams

Heuristics and Biases

"Judgments are all based on data of limited validity, which are processed according to heuristic rules" (Tversky & Kahneman, 1974, p.1124). Research on probability heuristics began in earnest in the 1970's and has been growing in popularity ever since. Two researchers, namely Amos Tversky and Daniel Kahnemam, have been especially prominent in this field. They established a cognitive understanding for common human errors in judgement and decision making that arise from heuristics and biases (Kahneman & Tversky, 1973; Kahneman, Slovic & Tversky, 1982; Tversky & Kahneman, 1974; Tversky & Kahneman, 1982, 1983). Together they developed Prospect Theory (Kahneman & Tversky, 1979), a behavioural economic theory that describes how people make choices between alternatives involving risk, where the probabilities of outcomes are known. They postulate that people base their decisions on the would-be losses and gains instead of the final outcome and use certain heuristics when making their evaluations. This theory was developed further in 1992 to overcome violations of stochastic dominance, which is when one gamble, or prospect, is ranked as superior to another by a large group of decision makers; making it a more accurate predictor of human judgement and decision making (Tversky & Kahneman, 1992). In 2002 Kahneman was acknowledged for his valuable contribution to our understanding of how human beings' reason under uncertainty, by being awarded the Nobel Memorial Prize in Economic Sciences for Prospect Theory in 2002[1].

[1]Amos Tversky was deceased by this time.

The term ‘heuristic’; however, has been around much longer than the 1970’s. This word emerged in the early nineteenth century from the Greek word *heuriskein* which means “find” (Online Etymology Dictionary, 2019). Originally heuristic meant enabling a person to discover or learn something for him or herself. This experienced-based technique for drawing conclusions and making decisions has become a seminal area in psychological research over the past half a century with many different judgement and decision heuristics emerging in the literature. The definitions of these heuristics, combined with the studies on them, suggest that people often rely on cognitive shortcuts when making judgements. The classic understanding of heuristics is that they save on effort but can also come at the cost of accuracy (Gigerenzer & Gaissmaier, 2011). In other words, heuristics offer a trade-off between cognitive efficiency and accuracy. When people need to make a difficult decision or judgement, they use a guideline that is simple and easy to access. Heuristics are cognitive shortcuts that people employ when a task or decision becomes too difficult. They are strategies or rules of thumb that reduce complex problems into simpler judgemental ones (Shiloh, Salton, & Sharabi, 2002). People normally do not have adequate formal models for computing the probabilities of uncertain events, thus intuitive judgement is often the only practical method for assessing uncertainty (Kahneman & Tversky, 1983). People tend to use heuristics when dealing with uncertainty; by relying on schemas to make guesses they are likely to arrive at the correct solution. “When faced with the difficult task of judging probability or frequency, people employ a limited number of heuristics which reduce these judgements to simpler ones” (Tversky & Kahneman, 1973, p. 207). Heuristics are quick and efficient in terms of time invested in making a decision. Usually they are well adapted to handle a wide range of problems (Kahneman & Frederick, 2005); however, if applied too broadly these same heuristics can bias judgements and decisions. They aid people to estimate the probability, or likelihood, of an event occurring but they do not guarantee accuracy (Evans, 2010). In other words, heuristics assist people to make judgements faster, but at the expense of occasional mistakes. The key word here is occasional; people make erroneous judgements sometimes, however continue to use heuristics in risky or uncertain situations because for the majority of the time the heuristics serve the task well. In fact, people either fail to realise the limitations of the heuristic, or they fail to realise they are using a heuristic in the first place because heuristics operate very quickly and automatically. If as Devine and Monteith (1999) describe, heuristic processes occur without intention, effort or awareness, and without interfering with other concurrent cognitive processes then it is possible that people are unaware of using them when solving problems.

Some researchers believe that heuristics are based on previous experience and knowledge and are considered to be serious sources of error due to their overuse or inappropriate application (Tversky & Kahneman, 1973). A prominent feature of judgement and decision-making research is the emphasis placed on errors – violations of rational norms – also known as biases. The normative theory of human judgement and choice plays a critical role in judgement under uncertainty because it provides a rational model to which participants' behaviour can be compared. In other words, these logical norms inform the null hypotheses of research studies. Although errors in judgement and choice are often interesting in themselves, this is not the reason why psychologists pay such close attention to them. It is through studying the biases that researchers are able to identify which heuristics were used to make certain judgements and decisions. "A bias is a systematic tendency to consider irrelevant factors and ignore relevant ones when solving a problem" (Moutier & Houdé, 2003, p. 186). A commonly adopted strategy for research in this field is to identify a factor that should generally *not* affect judgement or decision, and then design an experiment in which it does (Kahneman, 1991). Inducing and studying biases allows psychologists to determine what normal judgement and decision performance is, and it also accomplishes the methodological objective of rejecting a credible null hypothesis.

The Representativeness Heuristic. Determinants of representativeness are similarity and randomness. A sample is defined as *representative* if it is similar in essential properties to its parent population and reflects the salient features of the process by which it is generated. For example, how representative of a bird is a robin compared with a penguin? "A will be judged more probable than B whenever A appears more representative than B" (Kahneman & Tversky, 1972, p. 26). "In the same manner, a person seems representative of a social group if his or her personality resembles the stereotypical member of that group" (Tversky & Kahneman, 1983, p. 296). Thus, the ordering of events by their subjective probabilities coincides with their ordering by representativeness. A randomly selected sample should appear to be random in order for people to judge it as representative. Often people interpret coincidences as having more meaning than what they actually do, because they do not look random enough to be explained by chance. Take for instance a normal coin with one head (H) and one tail (T), if you were to toss it six times which of these two outcomes would you believe was more likely: T H H T H T or H H H T T T? Most people would guess that T H H T H T would be the most likely outcome of the two possibilities because it appears to be the most random, and a coin toss should have a random outcome (Teigen, 2004). However, both sequences are equally as likely to occur because the flip of a coin has a 50% chance of landing heads up or tails up. Representativeness is when a sample is considered more representative if it is similar in important characteristics to the population, or prototype, from which it was selected. People tend to ignore important information such as sample size and base rate in favour of representativeness, and also confuse independent and dependent events. For example, previous coin tosses cannot affect future ones, and thus previous tosses should not be taken into consideration when guessing probabilities of future tosses.

Representativeness is not always akin to similarity. Sometimes it can reflect a correlational and causal relationship amongst people's beliefs (Kahneman & Tversky, 1972). For example, a particular act such as stealing is representative of a particular person because we afford that person a specific disposition to commit the act, not because the act resembles the person. Representativeness also tends to covary with frequency. Common or frequent events are usually more representative than unusual or rare events. The representative summer day is warm and sunny, or the representative height of an adult female is 5 feet 4 inches. However, there are instances where a highly specific outcome can be representative but infrequent. For example, it might be representative for a female university student "to weigh between 124 and 125 pounds", but there are more "women who weigh more than 135 pounds" than "women who weigh between 124 and 125 pounds" (Tversky & Kahneman, 1983, p. 296). Thus "124-125 pounds" was judged as representative but less frequent than "above 135 pounds". An attribute can be representative of a class if it is very diagnostic. For example, it is more representative for a Hollywood actress to have been "divorced more than four times", than to "vote Democratic". Multiple divorces amongst Hollywood actresses feeds into a stereotype that divorce rates are higher amongst Hollywood actresses than among other women. An unrepresentative illustration of a category can be representative of a superordinate category. For example, rice is an unrepresentative vegetable, but it is representative of food. "Representativeness is nonextensional: It is not determined by frequency, and it is not bound by class inclusion. Consequently, the test of the conjunction rule in probability judgments offers the sharpest contrast between the extensional logic of probability theory and the psychological principles of representativeness" (Tversky & Kahneman, 1983, p. 296).

The belief that random looking outcomes are more likely than orderly outcomes is known as the *representativeness heuristic*. People often use this heuristic when they judge a sample to be more likely if it is similar to the population from which it was drawn (Kahneman & Frederick, 2002; Kahneman & Tversky, 1972). Although useful in many situations, the representativeness heuristic is extremely persuasive and can encourage errors when making more complex decisions. The representativeness heuristic is an example of a heuristic that often results in misconceptions of chance. It occurs when people form generalisations from highly limited information (Shiloh, Salton, & Sharabi, 2002). More specifically, when people consider a sample more representative if it is similar in important characteristics to the population – or prototype – from which it was selected. The stereotype or prototype of the population from which the sample is selected is often so compelling that people ignore important statistical information that otherwise should be have been considered. For example, sample size and base rate are two kinds of useful statistical information that are often overlooked in representativeness judgements.

Sample size and representativeness. Kahneman and Tversky proposed that that people assign probabilities to events so that "the more representative events are assigned higher probabilities, and equally representative events are assigned equal probabilities" (1972, p. 437). When making representativeness judgements people often fail to pay attention to the sample size due to the persuasiveness of the representation. A large sample is statistically more likely to reflect the true proportions in a population, while a small sample is more likely to reflect the extreme deviations from the population (Teigen, 2004). Consider the following example (Kahneman & Tversky, 1972; the correct answers are asterisked):

> *A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies born are boys. The exact percentage of baby boys, however, varies from day to day. Sometimes it may be higher than 50%, sometimes lower.*

For a period of one year, each hospital recorded days on which (more/less) than 60% of the babies born were boys. Which hospital do you think recorded more such days?

	(More than 60%)	*(Less than 60%)*
The larger hospital	*(12)*	*(9)**
The smaller hospital	*(10)**	*(11)*
About the same (i.e., within 5% of each other)	*(28)*	*(25)*

The researchers hypothesised that if people had any insight into the role of sample size, they should find it easy to select the correct answer. However, if they judge equally representative outcomes as equally likely, they should show no preference for the correct answer. When they tested this theory, they found that the latter was clearly the case. People opted for the "same" answer showing no preference for the correct answer. "The notion that sampling variance decreases in proportion to sample size is apparently not part of man's repertoire of intuitions" (Kahneman & Tversky, 1972, p. 444). People make these sampling biases often in everyday life, for example people are willing to take seriously a result expressed as a percentage without concern for the number of observations, which may be ridiculously small. Alternatively, people often remain sceptical in the face of solid information from a large sample. The researchers conclude that people can be taught the rule of sampling; however, work by Tversky and Kahneman (1971) shows a strong tendency to underestimate the impact of sample size, even when the subjects have knowledge of the correct rule and extensive statistical training.

Base rate and representativeness. Just as people neglect the sample size, so do they ignore the base rate when making representativeness judgements. Again, this is due to the persuasiveness of the heuristic. The base rate is the amount of times an item occurs in the population, and people tend to rely on representativeness when they are asked to judge category membership (Teigen, 2004). Kahneman and Tversky (1973) devised a study that tested the "hypothesis that intuitive predictions are dominated by representativeness and are relatively insensitive to prior probabilities" (p.241). Subjects were presented with the following cover story:

> *A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.*
>
> *The same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions. You will be paid a bonus to the extent that your estimates come close to those of the expert panel.*

There were two conditions; the low engineer group (e.g., 30 engineers and 70 lawyers) and the high engineer group (e.g., 70 engineers and 30 lawyers). All subjects were presented with five descriptions, one of which read:

> *Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.*
>
> *The probability that Jack is one of the 30 engineers in the sample of 100 is _____%.*

Following the five descriptions, the subjects were shown the null description:

Suppose now that you are given no information whatsoever about an individual chosen at random from the sample.

The probability that this man is one of the 30 engineers in the sample of 100 is ___%.

The researchers found that the average value was 50% for the low-engineer group, and 55% for the high-engineer group, and although the effect of prior probability was slight it was significant ($p < .01$). However, the researchers also discovered that the only condition that provided the correct response (e.g., 70% for high-engineer group and 30% for low-engineer group) was the null description. All five other conditions gave incorrect responses. Thus, the researchers concluded that explicit manipulation of the prior distribution had minimal effect on subjective probability. Participants applied their knowledge of the prior only when they were given no specific evidence, and prior probabilities were largely ignored when individuating information was made available. Indeed, even giving the participants worthless information resulted in errors. The example read:

Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

This description was constructed to be completely uninformative with regards to Dick's profession and participants responded with 50% for both the low- and high-engineer groups. When no specific evidence is given, in other words, by removing the context, it seems that prior probabilities are properly utilised. However, by including evidence, even worthless evidence, those prior probabilities are ignored. These findings highlight that people focus almost exclusively on representativeness when judging category membership, and by doing so commit the base-rate fallacy. In other words, the important statistical information found in the base rate in underestimated, while representativeness is overemphasised; resulting in a failure to identify that one category is more common than the other.

In 2007, Barbey and Sloman published a review article on the phenomenon of base-rate neglect. They attempted to explain human rationality, and how people make judgements under conditions of uncertainty by unpacking the perspectives found in the literature and evaluating how well they explained existing data and their conceptual coherence. Their empirical review supported five main conclusions, all of which suggested that people do not have an inductive reasoning mechanism designed to process natural frequencies, instead they rely on cues to the set structure of the problem for accurate reasoning. In fact, partitioning the data into nested sets facilitated Bayesian reasoning regardless of whether natural frequencies or single-event probabilities were used. The authors found solid evidence supporting the nested set hypothesis. In other words "the mind embodies a domain general capacity to perform elementary set operations and these operations can be induced by cues to the set structure of the problem to facilitate reasoning in any context where people tend to rely on associative rather than extensional, rule-based processes" (Barbey & Sloman, 2007, p. 252). Errors were reduced when the question was presented in such a way that it made the participants use the sample of category instances to give an answer, or when the sample of category instances were represented in a nested set structure, which organised the data into a manner needed to solve the Bayesian computation.

The Conjunction Fallacy

Following the early work of Kahneman and Tversky in the 70's and 80's, a vast amount of empirical data has emerged all casting doubt on the extent to which human reasoning is logical (see Evans & Over, 1997; Kahneman, 2011). An example of this is how poorly people perform on intuitive probability judgement tasks (Tversky & Kahneman, 1974, 1983). A realization from all of these studies is that "adult subjects are able to correctly carry out certain probabilistic reasoning tasks—which is a testimony to their logical inductive abilities—while still exhibiting massive reasoning biases in other contexts" (Moutier & Houdé, 2003, p.186). The conjunction fallacy is one of the classical intuitive probability judgement biases that affects inductive reasoning. Inductive reasoning is when specific observations lead to a general or probable conclusion, compared to deductive reasoning which is when general rules lead to a true and specific conclusion. An example of inductive logic is, "The coin I pulled from the bag is a penny". Even though all the premises are true in this statement, inductive reasoning allows for the conclusion to be false. Tversky and Kahneman (1974, 1983) were the first to demonstrate the conjunction fallacy using both frequency and probability judgements.

The conjunction fallacy is perhaps the most notorious of dual-process reasoning errors. When a person neglects to apply the laws of conjunction probability and rates a conjunction of events occurring as more likely than either of the constituent events occurring singularly, then they have committed a conjunction fallacy. In other words, they assume that multiple specific conditions are more probable than a single general one. Often seduced by compelling heuristics the reasoner commits this bias despite it being statistically impossible. Take these two scenarios for instance, 1) a car accident occurring on a perfectly straight highway, and 2) a car accident occurring on a perfectly straight highway with a grease spot. The second statement may be judged to be more likely to occur given the grease spot as this gives rise to the stereotype of a slippery road. However, the second statement is merely an extension of the first. In both scenarios the car accident occurred on a perfectly straight highway. Another example would be the probability of having three daughters and one son may appear higher, intuitively, than the probability of having only three daughters (Fiedler, 1988). Tversky and Kahneman attribute the conjunction fallacy to the representativeness heuristic (1983). They believe that people derive their probability estimates by application of this heuristic, and therein lies the problem because this then leads to biases. The conjunction of two events combines one representative (likely) event with another unrepresentative (unlikely) event, and this conjunction is usually perceived as more representative (likely) than the unrepresentative event on its own. These judgements via representativeness give rise to erroneous probability judgements, in other words, the conjunction is judged more probable than its unrepresentative (unlikely) component. “If probability estimates were based on representativeness, it would seem reasonable to assume that the more representative the larger component, the more representative the conjunction will be judged to be and the greater will be the extent of the conjunction fallacy” (Fisk, 2002, p. 432). In 1977, Tversky proposed a model where similarity is a continuous function of both similar and dissimilar events. Tversky and Kahneman (1983) have suggested that this model might provide the basis for judgements under representativeness, such that the perceived representativeness of the conjunction depends on the perceived representativeness of both the unlikely (unrepresentative) component and the likely (representative) component. Hence, the more likely (representative) the larger component is perceived to be, the more likely (representative) the conjunction will be judged to be.

In a pioneering series of studies, Tversky and Kahneman (1983) demonstrated that, when evaluating the likelihood of the conjunction of two events, most people made these systematic errors. Among the best known examples of this type of logical violation is the ‘Linda problem’.

The Linda Problem. The representativeness heuristic involves using a stereotype or prototype when making a judgement instead of logic, and this can lead to errors in judgement. To test this idea Tversky and Kahneman (1983) developed a task affectionately known today as the “Linda problem”. Their earliest studies on the conjunction rule began in

1974 and used occupation and political affiliation as target attributes to be predicted singly or in conjunction from brief personality sketches. In 1983 they replicated and extended this work using the following personality sketches of two fictitious people, Linda and Bill (see Figure 1. 2), followed by a list of occupations and hobbies associated with each of them. The task involved reading a personality sketch of Linda and Bill followed by a list of eight statements. These statements were comprised of hobbies and occupations, some more closely associated to the prototype of Linda and Bill than others. Participants were required to read the personality description and then to rank the eight statements in order of the degree to which they resembled a typical member of that class. They gave a ranking of "1" for the most likely option and a ranking of "8" for the least likely option. The description of Bill was constructed to be representative of an accountant (A) and unrepresentative of a person who plays jazz for a hobby (J), while the description of Linda was representative of an active feminist (F) and unrepresentative of a bank teller (T).

Read the following paragraph:	Read the following paragraph:
Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations.	Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.
Now rank the following options in terms of the probability of their describing Linda. Give a ranking of 1 to the most likely option and a ranking of 8 to the least likely option:	Now rank the following options in terms of the probability of their describing Bill. Give a ranking of 1 to the most likely option and a ranking of 8 to the least likely option:
___Linda is a teacher in elementary school.	___Bill is a physician who plays poker for a hobby.
___Linda works in a bookstore and takes yoga classes.	___Bill is an architect.
___Linda is active in the feminist movement. (F)	___Bill is an accountant. (A)
___Linda is a psychiatric social worker.	___Bill plays jazz for a hobby. (J)
___Linda is a member of the League of Women Voters.	___Bill surfs for a hobby.
___Linda is a bank teller. (T)	___Bill is a reporter.
___Linda is an insurance salesperson.	___Bill is an accountant who plays jazz for a hobby. (A&J)
___Linda is a bank teller and is active in the feminist movement. (F&T)	___Bill climbs mountains for a hobby.

Figure 1. 2. Demonstration of the Linda and Bill problems designed to investigate the influence of the representativeness heuristic on conjunction probability judgements (Tversky & Kahneman, 1983)

Would you agree that Linda is more likely to be a bank teller than she is to be a bank teller *and* active in the feminist movement? Most people would not agree with this statement because Linda more closely resembles a feminist than a bank teller (Tversky & Kahneman, 1983). Linda is described as a stereotypical feminist, thus our instinct would be to label her as one even if this meant choosing a conjunction of two attributes (bank teller and feminist) as more likely than one attribute (bank teller). However, it is logically impossible for two attributes to be more likely than one. The conjunction fallacy occurs because people naturally form a heuristic impression of a believable answer using their initial intuitive judgement. In the Linda problem, this is a portrayal of a feminist because Linda has many traits and hobbies representative of a woman actively involved in the feminist movement. Indeed, the results showed that the majority of participants correctly ranked the representative single event as more likely than both the unrepresentative single event and the conjunctive events. However, more surprisingly to the authors was the fact that the participants incorrectly ranked the conjunctive events as more likely than the unrepresentative single event, and in doing so committed the conjunction fallacy (87% responded A > A&J > J for Bill; 85% responded F > T&F > T for Linda). According to the conjunction rule, the conclusion T&F > T is mathematically impossible. The conjunction law of probabilities states that the probability of the conjunction of two events can never be larger than the probability of either of its constituent events; $P(\text{T\&F}) \leq P(\text{T})$ or $P(\text{F})$. In the Linda problem, the conjunction of the two events (bank teller and feminist) cannot occur more often than either event by itself (i.e., being either a bank teller, or an active feminist). When performing conjunction probability tasks most people mistakenly believe that the co-occurrence of two events (Linda is a bank teller AND a feminist) are more likely than the occurrence of one of those events (Linda is a bank teller), and in doing so they commit the conjunction fallacy. "This is called the conjunction fallacy because the class of bank tellers includes the class of bank tellers who are feminists; therefore, being a bank teller who is a feminist cannot be more probable than being a bank teller" (Reyna, 1991, p. 318). If representativeness depends on both common and distinctive features (Tversky, 1977), it would be heightened by the addition of shared features. In this example, the addition of active feminist to the occupation of bank teller improved the match between Linda's personality sketch and her current activities (Tversky & Kahneman, 1983). Tversky and Kahneman (1983) believe that erroneous conjunction judgements of probabilities are the result of the representativeness heuristic. Instead of deliberating and using logic to find the solution, people simply rely on the heuristic to make a fast, intuitive decision that leads to the representative – and often incorrect – answer. These findings have contributed to the widespread belief that traditional logical or probabilistic considerations play little role in our reasoning (Gigerenzer, 1996). In an attempt to further understand why people were so heavily influenced by the representativeness heuristic when making probability judgements under uncertainty, Tversky and Kahneman (1983) elaborated upon this research by introducing three new conditions to the Linda problem: the indirect, direct-subtle, and direct-transparent versions. All three versions used the same personality descriptions as in the original study (i.e., Linda and Bill); the differences are in the statements presented afterwards.

The *indirect* version required one group of participants to judge the probability of the conjunction (T&F), whereas another group judged its constituents (T, F). There was no requirement to compare the conjunction with the constituents, thus testing whether probability judgments conformed to the conjunction rule. The *direct-subtle* version required all participants to compare the conjunction to its less representative constituent (T, T&F). The inclusion relation of the two events was not emphasized, thus testing whether people would take advantage of an opportunity to compare the critical events. The *direct-transparent* version required participants to evaluate the probabilities of the conjunction and its constituents (T&F, T, F). The inclusion relation between these statements was explicitly highlighted, which tested whether people would conform to the rules of logic of probability when they were led to compare the critical events. These three variants were tested on three groups of participants: statistically naïve (no statistical knowledge or background), informed (familiar with basic probability theory), and sophisticated (doctoral students who had taken advanced statistics and probability courses). To the researchers' surprise, all three groups ranked the conjunction higher than its less likely constituents, there were no consistent differences found between the rankings of the direct and indirect versions of the test, the overall rate of violations of the conjunction rule was 88% for direct tests, and there was no effect of statistical sophistication in either the direct or indirect tests. The researchers had expected that participants would make less conjunction errors in the direct tests compared to the indirect tests, and had also expected that even naive participants would be able to apply the conjunction rule. This was clearly not the case as in both direct and indirect versions of the test all the participants, including the statistically sophisticated ones, committed the conjunction fallacy.

These surprising results motivated the researchers to embark on a "series of increasingly desperate manipulations designed to induce subjects to obey the conjunction rule" (Tversky & Kahneman, 1983, p. 299). First, they presented the personality description of Linda to a group of undergraduate participants and asked them to check which of these two alternatives was more probable "Linda is a bank teller (T)" or "Linda is a bank teller and is active in the feminist movement (T&F)". Overall, 85% of respondents indicated that T&F was more probable than T, in a "flagrant violation of the conjunction rule" (Tversky & Kahneman, 1983, p. 299). The authors then decided to investigate whether there was a misinterpretation of the wording in the task. If participants believed the wording to be too trivial, they might not have taken the statement literally. In other words, "Linda is a bank teller" might have been interpreted as "Linda is a bank teller and is *not* a feminist". If this were in fact the case, then ranking T&F higher than T would not have been a conjunction error. They tested this idea by asking a group of participants to judge the probability of T and of T&F on a 9-point scale ranging from 1 (extremely unlikely) to 9 (extremely likely). By rating the probabilities this way – when one of the events includes the other – there was no reason for respondents to interpret T as T&notF. The results obtained were the same as before with 82% of participants ranking T&F as more likely than T. To examine whether people are able to identify the validity of the conjunction rule, even though they are unable to apply it spontaneously, Tversky and Kahneman presented the Linda description to a further group of participants and asked them to state which of two arguments they found most convincing. Argument 1: "Linda is more likely to be a bank teller than she is to be a feminist bank teller, because every feminist bank teller is a bank teller, but some women bank tellers are not feminists, and Linda could be one of them (T)", or argument 2: "Linda is more likely to be a feminist bank teller than she is likely to be a bank teller, because she resembles an active feminist more than she resembles a bank teller (T&F)". Although the results showed an improvement, the majority of participants (65%) still committed the conjunction fallacy despite a deliberate attempt at inducing a reflective mood. Finally, the authors attempted to highlight the inclusive nature of the event T by conveying it as a disjunction. The conjunction rule can also be expressed as a disjunction rule: $P(\text{A or B}) \geq P(\text{B})$. Once again, the Linda description was presented to a group of participants who then had to rate the probability statements on a 9-point scale ranging from 1 (extremely unlikely) to 9 (extremely likely); however, the statement T was replaced by "Linda is a bank teller whether or not she is active in the feminist movement (T*)". This adaptation was designed to highlight the inclusive nature of T&F in T. Overall, 57% of participants violated the conjunction rule because they failed to draw extensional inferences from the phrase "whether or not". The authors then asked participants to rank T, T*, and T&F on a 9-point scale and found that people believed T < T* < T&F to be true. They surmised that because Linda is representative of an active feminist, the mere mention of the word "feminist" causes T* to be rated as more likely than T, despite them being extensionally equivalent. And a definite commitment to "feminist" makes the probability of T&F even higher.

The high rates of conjunction fallacies perplexed the authors and they continued searching for reasons for this and ways to improve this phenomenon. They managed to achieve lower rates of the conjunction fallacy by asking participants to bet on T or on T&F in the transparent version of the task. Subjects were given Linda's description followed by the following instructions: "If you could win $10 by betting on an event, which of the following would you choose to bet on? (Check one)". Conjunction fallacies were committed by "only" 56% of individuals, which is still high but substantially lower than the typical value circa 80%. The lowest rates of conjunction fallacies achieved was by testing a group of statistically savvy students. Previous samples had used statistically naïve undergraduates, which prompted the authors to recruit graduate students who had undertaken several statistics courses. They completed the rating-scale version of the direct test of the conjunction rule for the Linda problem. For the first time ever, they found that only 36% committed the conjunction fallacy. This finding is still rather high for the a group of intelligent and statistically sophisticated respondents, however it seemed that statistical sophistication produced a majority who conformed to the conjunction rule (in a transparent test).

The findings discussed above caused Tversky and Kahneman (1983) to question whether the conjunction fallacy holds for other areas of judgement, and whether or not expertise in the relevant subject matter protected against the conjunction fallacy? They found that the phenomenon of committing the conjunction fallacy was extremely robust and this applied across domains such as medicine, gambling, political forecasting, and law. The authors tested over 3000 participants and found widespread violations of the extension rule by naïve and sophisticated participants in both the indirect and direct versions of the problem.

Other Conjunction Fallacy Tasks. The natural assessments of representativeness and availability do not conform to the extensional logic of probability theory. In other words, a conjunction judgement can be seen as more representative than one of its constituents, and in instances of recalling categories, a specific category can be easier to retrieve than a more inclusive category (Tversky & Kahneman, 1983). The Linda problem formed the foundation for many conjunction probability judgement tasks to follow. The most popular adaptations have been designed using the same format of providing the participant with a descriptive sketch of a fictitious person and then asking the participant to rank a number of statements in order of likelihood from most to least, or to estimate their likelihood of occurring using a Likert scale. In addition to probabilities, another example of the conjunction fallacy can be found in judgements of frequency, such as the seven-letter word test. Tversky and Kahneman devised the seven-letter word task in 1983, where they gave participants 60 seconds to list seven-letter words of a specified form and found that their student participant sample produced more words of the form _ _ _ _ *i n* g than the form _ _ _ _ _ *n* _, despite the latter class including the former. The researchers found that the average number of words produced in the two conditions were 6.4 and 2.9, respectively. "In this test of availability, the increased efficacy of memory search suffices to offset the reduced extension of the target class" (Tversky & Kahneman, 1983, p.295). Similar results were obtained for the comparison of words in this form _ _ _ _ _ *l y*, compared to this form _ _ _ _ _ *l* _, with median estimates of 8.8 and 4.4, respectively.

A similar example of the seven-letter word task can be seen in Kahneman and Tversky (1996). They displayed the seven-letter word task, which required participants to estimate the number of seven-letter words conforming to the following form: _ _ _ _ _ *n* _ in a piece of text approximately 2000 words, or 4 pages long. Following this, the same participants estimated the number of seven-letter words conforming to a new form:

_ _ _ _ *i n g* in a piece of text 4 pages long. Due to the availability heuristic, it is easier to think of words ending in “i n g” than “_ n _” (despite them being the same), thus participants will overestimate the frequency of the former, and in doing so will commit a conjunction fallacy. Kahneman and Tversky (1996) reported that individuals estimated words ending in “ing” nearly three times higher than those ending in “n” in the next-to-last position. The researcher’s interpretation of these findings was that people are capable of making correct probability judgements; however, due to significant judgmental heuristics they fall prey to systematic judgmental biases.

The Linda and seven-letter words tasks are both inductive, conjunction fallacy tasks but differ in presentation and wording. The Linda problem deals with probability judgements while the seven-letter word task deals with frequency judgements. However, they both provoke similar judgemental heuristics (i.e., representativeness and availability) that lead to committing the conjunction fallacy. Both tasks rely on the same underlying logic and require people to apply the inclusion rule and the conjunction probability rule in order to solve them correctly, or logically. According to Moutier and Houdé (2003), to provide a correct answer an individual would be required to 1) inhibit the misleading scheme (i.e., the judgemental heuristic) and 2) activate the relevant scheme (i.e., inclusion rule and conjunction probability rule).

What we know of the Conjunction Fallacy. The conjunction fallacy can be found in a multitude of settings which is why it is so important for researchers to gain a better understanding of why and how we make these errors in reasoning. Conjunction fallacies have been observed in the case of probabilistic reasoning in legal decision making. For example, Wojciechowski and Pothos (2018) found that people without a legal background were committing the conjunction fallacy by rating the probability that a suspect committed two crimes higher than the probability that they committed one of the constituent crimes.

Interestingly, these findings did not convey to judges and attorneys with a legal background. In the context of profiling victims of homicide, Dearden (2018) found that people overestimated the likelihood that certain situational or demographic characteristics would occur together, thus committing the conjunction fallacy. For example, "the overwhelming number of victims, 183, were Black as were the suspects, 66. More than half the victims were between the ages 18 and 29, and 181 were male. Handguns continued to be the choice of murder weapon, in 149 killings" (Hermann, 2012). Participants were swayed by the profiles or stereotypes of the victims causing them to consider these demographics together, instead of separately. Each of these statistics is independent. Being Black has no relationship to being male. However, consistent with the conjunction fallacy, many participants overestimated the likelihood that these demographics occurred together, and some to a "mathematically impossible degree" (Dearden, 2018, p. 187).

In 2003, Medin, Coley, Storms, and Hayes demonstrated a parallel conjunction fallacy in category-based induction. This effect was demonstrated for both property reinforcement arguments and causal arguments. Causal arguments are causally related, such as: "Grain has property *M4*. Therefore, mice have property *M4*." While property reinforcement arguments share categories, for example: "Andean people have property *X12*. Therefore, Himalayan people have property *X12*." Their results showed that across a number of sets of arguments the mean ratings of inductive strength (on a 9-point scale) for two-conclusion category arguments (e.g., "Therefore, mice and owls have property *M4*" and "Therefore, Himalayan people and Alpine people have property *X12*") were greater than single-conclusion category arguments. This is normatively contradictory, as the probability of a conjunction cannot be greater than the probability of its constituents; thus, causing a conjunction fallacy. The differences between causal and categorical materials elude to the fact that the conjunction fallacy might be based on different kinds of knowledge about the relationships among categories. This idea led Feeney, Shafto and Dunning (2007) to investigate further. They replicated Medin et al.'s (2003) finding of a conjunction fallacy in category-based induction. Also using both causal and categorical (i.e., what Medin et al. referred to as property) arguments to test the conjunction fallacy, but this time they took cognitive ability into account. Their results showed that people are more susceptible to the causal version of the fallacy than the category-reinforcement version. They also found that high ability participants were less susceptible to the conjunction fallacy in the category reinforcement condition; however, in contrast, there was no relationship between ability and susceptibility in the causal condition. The researchers stressed the importance of including both how individual differences in cognitive ability and different types of prior knowledge affect reasoning, when forming a complete theory of reasoning. Stanovich and West (2008) explored the effects of cognitive ability on thinking biases. They found contradictory results showing that both highly intelligent and less intelligent people (as measured by SAT scores) committed the conjunction fallacy. However, more surprisingly, they discovered that the interaction between fallacies and cognitive ability was in the opposite direction. Indeed, the group with higher SAT scores was more susceptible to the conjunction fallacy.

Brotherton and French (2014) investigated whether a relationship existed between the belief in, and endorsement of, conspiracy theories and the conjunction fallacy. It is thought that anomalistic beliefs are accounted for (at least in part) by the representativeness heuristic: the automatic assumption that 'like causes (or is caused by) like' (Kahneman & Tversky, 1972, Teigen, 2004). It makes sense that a bias in judgement that overestimates the likelihood of co-occurring events (Kahneman & Tversky, 1972) would correlate with paranormal beliefs, as paranormal believers have a "biased conception of randomness" (Brotherton & French, 2014, p. 246) and base their judgements on certain coincidences (i.e., subjective representativeness) rather than on objective probabilistic. Across a series of two experiments, the authors found that both paranormal believers and conspiracy believers made more conjunction errors than nonbelievers. The stronger people endorsed conspiracy theories the more conjunction errors they committed. The effects were not limited to conspiratorial items, which suggests that individual differences in susceptibility to the conjunction fallacy are domain-general. Surprisingly, participants in these studies committed conjunction fallacies 33.3% and 48% of the time, which is still high but nowhere near as high as other studies in the field (e.g., Tversky and Kahneman, 1983, found 85%). It is unclear from these results whether susceptibility to the conjunction fallacy causes or conversely is caused by endorsement of conspiracy theories. Although the authors acknowledge that it might be a reciprocal relationship, they do mention that susceptibility to commit the conjunction fallacy is domain-general, which suggests that some people are inherently more susceptible to the conjunction fallacy than others.

In 2009, Crisp and Feeney conducted research in an attempt to bridge a gap in our understanding of causal conjunction fallacies. A causal conjunction is when the conjunction includes a possible cause and an outcome, and a fallacy is committed because the strength of the casual link biases the probability judgement. Consider these two statements for example taken from Sides, Osherson, Bonini and Viale (2002), "Smoking will decrease by 15%" (A) and "The government will increase tax by 1$ per pack and smoking will decrease by 15%" (B). By mentioning the increase in tax and thus the increase in price of cigarettes, they have strengthened the causal relationship between the two conditions. Hence making it more appealing to choose the conjunction statement. Crisp and Feeney (2009) predicted that the greater the strength of the casual connection between constituent events would directly affect the magnitude of the causal conjunction fallacy and tested this by using weak, strong, and unrelated causal events. Findings from the first experiment showed that rates of violations were higher for strongly-related than for weakly-related conjunctions, which in turn produced higher fallacy rates than unrelated conjunctions. The second experiment included a cognitive load condition (i.e., to suppress deliberative thinking) as a means to better understand the thinking systems used when making causal conjunction judgements. The results showed that the dot memory task increased flawed responding for strongly related events but had no effect on the fallacy rate for weakly related events. In other words, strongly related conjunctions cue a stronger heuristic output than weak conjunctions, but when under cognitive load people lack the mental resources to inhibit the compelling heuristic response. Thus, causal knowledge has a crucial influence on probabilistic reasoning as varying causal strength also influences the strength of the cued heuristic response.

Methods of improving Conjunction Violations. Bakhti (2018) found that using priming methods affected the performance on conjunction judgements. Participants were allocated into one of three groups: religious priming, reflective priming and neutral. Priming was presented through the scrambled sentence task. Religious priming involved unscrambling sentences made from religious words (e.g., pray for the poor), reflective priming was unscrambling analytical words (e.g., I think all day) and the neutral condition included unscrambling neutral words (e.g., he finished it yesterday). The results showed that people who received religious priming committed more conjunction fallacies than those who received reflective priming. However, there was no difference found between religious and neutral priming, therefore indicating that reflective priming reduced conjunction fallacies; while religious priming had no effect.

A recent Brazillian study showed that the conjunction fallacy was a robust phenomenon that could, however, be minimized by converting all the probability alternatives into conjunctions. Usually the list of alternatives shown to participants include a number of single event and one conjunction that contains two of the listed single events. Campos, Lincoln, Neves, Correia and Soares (2013) conducted a series of three experiments that firstly, emphasized the resilience of the conjunction fallacy and secondly, showed what measures could be taken to minimize the effect on human judgement. In the first two studies, the authors demonstrated the robustness of the conjunction fallacy phenomenon by achieving levels of conjunction fallacies at 76% and 86% respectively. The third and final study attempted to minimize the effect of the conjunction fallacy by presenting the information differently. The authors discovered that by presenting all the statement alternatives as conjunctions, they were able to reduce the percentage of errors in judgement from an average of 80.4% down to only 17.2%.

Charness, Karni and Levin (2010) were not convinced that previous research on the conjunction fallacy was relevant to real-life economic behaviour, where people are able to discuss their options before deciding and often these decisions involve monetary incentives. Thus, they designed an experiment aimed to replicate the results of Tversky and Kahneman (1983) using the transparent test as their main experiment. Participants either made unassisted (e.g., individual) or assisted judgements (e.g., they were allowed to consult with either one or two other participants before giving their answer). And they were either offered small monetary incentives, or no incentives at all. The results revealed far lower rates of conjunction fallacies (e.g., 58%) compared to Tversky and Kahneman (e.g., 85%), and when incentives were introduced that rate dropped to 33%. Additionally, when communication between participants was allowed violation rates fell significantly, and this effect was most noticeable in groups of three, compared to groups of two, or even one person. These findings suggest that the rate of conjunction fallacies is in fact much lower than reported in Tversky and Kahneman (1983), and this error rate declines dramatically when subjects are given the opportunity to consult with others prior to making a judgement. Unsurprisingly, financial incentives for providing the correct answer were effective. In light of this evidence, the authors surmise that the "conjunction fallacy is a phenomenon of no serious concern for economics” (p. 555) as people rarely make decisions in isolation. They consult with others and also recognize that their choices have consequences. However, contradictory to these findings, Bonini, Tentori and Osherson (2004) found that most people do not spontaneously integrate the logic of conjunction into their assessments of chance. They designed experiments using a betting paradigm that discouraged misreading p as p-and-not-q and encouraged genuine conjunctive reading of p-and-q. Participants were instructed to divide 7 euros among 3 statements with the understanding that the assigned amount could be won if the statement comes true. Due to the fact that the conjunction p-and-q implies p, the value of a bet on p-and-q cannot exceed the value of a bet on p at the same stakes; however, frequent violations were observed. The authors rule out the possibility of miscommunication as an explanation, and instead attribute the fallacies to a failure to reconcile probability with logic.

Moutier and Houdé (2003) suggested that systematic judgemental biases, such as the conjunction fallacy, are due to inhibition failure and not statistical sophistication. To test this theory, they designed two types of training before completing the seven-letter word task. First, a strictly logical procedure in which participants were trained to use the conjunction rule, and second, both a logical and executive-emotional procedure in which participants were alerted to the bias through warnings such as “*we’re falling into a trap!*” and “so the goal here is (1) *not to fall into the trap*” (p. 192). The words “not to fall into the trap” evoked both an emotional response and also a tendency to avoid the biased answer. Results showed a significant effect of inhibition training; however, strictly logical training (e.g., without executive-emotional warning) had no effect on performance. This confirms that cognitive inhibition is key to correcting intuitive judgements. Interestingly, not all participants benefitted from the training, which raised the question of what accounts for the inter-individual variability of receptivity to inhibition training? This question led Cassotti and Moutier (2010) to investigate further. They reconfirmed that training based solely on logical explanations is in fact ineffective to overcome System 1 intuitive-heuristic reasoning and found that executive-emotional warnings enabled drastic changes in reasoning performance. Further, they found that participants who did not respond to the inhibition training mainly had less emotion-based learning ability as measured by the Iowa Gambling Task, whereas the opposite was true for those who did learn from the training. The authors surmised that in order to resist the bias, participants with lower emotional warning sensitivity needed greater inhibitory efficiency (e.g., more in-depth inhibition training) than those with higher emotional warning sensitivity. Houdé et al. (2000) explored closer, through the use of functional imaging (PET), what happens in the mind when the brain has to inhibit an intuitive response to activate logical reasoning. They focused on matching-bias tasks where participants completed a pretest, followed by inhibition training, and finally a posttest. The authors noticed a "striking biased-to-logical shift under matching-bias inhibition training" (p. 5). The PET results showed that brain activity shifted from the posterior part of the brain when performing the pretest, to the left prefrontal network when completing the posttest. This reflects the change in the participants' reasoning strategy as a specific consequence of the executive-inhibition training. This finding shows how the activated brain networks can change in the same subjects, carrying out the same task, depending on their ability to inhibit a misleading strategy.

A well renowned method for alleviating the conjunction fallacy is by presenting the information in frequencies as opposed to probabilities. Gigerenzer (1994) claimed that the conjunction fallacy disappears when formulated in terms of frequencies. In other words, by adding the following sentence after the description of Linda: "There are 100 people who fit the description above", the experimenter has managed to change the structure of the task from a probability problem to a mathematical or frequentist problem. And in doing so, has changed the way people think when solving the problem. "The conceptual distinction between single-event and frequency *representations* suffices to make this allegedly stable cognitive illusion largely disappear" (Gigerenzer, 1994, p. 144) Fiedler (1988) reported that the number of conjunction violations in the Linda problem dropped from 91% in the original, single-event representation to 22% in the frequency representation. While Reeves and Lockhart (1993) found 59% conjunction violations for probability problems and 35% for frequency problems. Hertwig and Gigerenzer (1994, cited in Gigerenzer, 1994, p. 144) asked participants to rank order single-events (T, T&F, and F) according to their likelihood of occurrence, and noticed a conjunction fallacy rate of 88%; however, when asked to estimate the frequency of T, T&F and F (e.g., "how many out of 200?") those violation rates dropped to 13% and 16% respectively across two frequency experiments. What is more, this effect is stable across different cognitive illusions, including base rate neglect. Gigerenzer (1994) concludes that we must pay attention to the kind of information representation that algorithms were designed to work upon. To discuss rationality simply in terms of good or bad algorithms is incomplete, because when information is represented in terms of frequencies rather than single-event probabilities, these apparent stable cognitive fallacies disappear. However, Kahneman and Tversky (1996) made an argument against the frequency hypothesis by pointing out that biases have been demonstrated with frequency judgements since the emergence of the heuristics and biases program. Research conducted by Sloman, Over, Slovak and Stibel (2003) compared the natural frequency hypothesis to the "nested-sets" hypothesis (i.e., "presenting the options as concrete classes made the inclusion relation between the two sets more transparent" p. 305). The authors found that the benefits of frequency in reducing the conjunction fallacy were eliminated by making the nested-set relationship obscure by spacing the conjunction and its constituent. They make the observation that nested-set relations are more general than frequency representations. While the nested-set hypothesis increases transparency in the task, thus increasing coherence of probability judgements; this is not the same as inducing a frequency frame. The nested sets only need to represent the subset relation between the conjunction and constituent, not the actual frequency of each. Interestingly, they also found that the way participants were asked to answer the tasks affected the number of conjunction fallacies committed. To be more specific, ranking responses produced more conjunction fallacies than rating responses. This could be due to the fact that ranking forces a choice between the critical statements. It does not offer the option to judge two alternatives as equally probable. Thus, some people may have committed the fallacy despite the conviction that the two options were equally likely. Ratings, on the other hand, allow people to assign equal probabilities to the two events.

Hertwig and Gigerenzer (1999) tested the hypothesis that between-subjects designs lessen the frequency of conjunction violations. They tested this hypothesis using the seven-letter word problem both in a within-subjects and between-subjects design. They found that the between-subjects groups fell prey to the conjunction fallacy by estimating that more words end in 'ing' than in '_n_'. However, the within-subjects group, who saw both alternatives, answered the questions correctly and only 26% committed the conjunction fallacy. This finding is not task specific as Birnbaum and Mellers (1983) discovered that the base rate fallacy also leads to different conclusions depending on whether a within-subjects or between-subjects design is used.

Alternative explanations for the Conjunction Fallacy. There exists a dispute in the literature on whether the conjunction fallacy is the product of people's *misunderstanding* of the problem, or whether it is a case of genuine *reasoning bias*. Over the years, researchers have offered alternative explanations for the occurrence of the conjunction fallacy. One such prevalent explanation is that of linguistic ambiguities. Camerer (1995) explains the high rate of violations of the conjunction principle in terms of linguistic conventions:

> ... some apparent biases might occur because the specific words used, or linguistic convention subjects assume the experimenter is following, convey more information than the experimenter intends. In other words, subjects may read between the lines. The potential linguistic problem is this: in the statement "Linda is a feminist bank teller," subjects might think that this statement "Linda is a bank teller" tacitly excludes feminists; they might think it actually means "Linda is a bank teller (and not feminist)." If subjects interpret the wording this way none of the statements are conjunctions of others and no probability rankings are wrong. (p. 598)

Hertwig, Benz and Krauss (2008) postulate that unintended interpretations of the connective *and* may account for conjunction violations. They state that the word 'and' is used in the task to invoke the logical connective ∧ whereas when presented experimentally it is being interpreted with natural language (e.g., English) as a conjunction *and*. Unlike the former, the latter *and* relays a wide range of relationships between conjuncts, and depending on which meaning of *and* is adopted "people may arrive at nearly opposite understandings of a sentence" (Hertwig, Benz & Krauss, 2008p. 741). There is; however, relevant evidence against this version of the misunderstanding hypothesis. Bar-Hillel and Neter (1993) found high percentages of the conjunction fallacy in contexts where the word 'and' does not appear. While Bonini, Tentori and Osherson (2004) used an explicit reminder of the conjunctive meaning of 'and' to discourage a disjunctive interpretation of 'and' (e.g., "both events must happen for you to win the money on this bet" p.204). However, most participants continued to commit the conjunction fallacy in this context. A control task run by the same authors confirmed that around 91% of the participants believed that event A was true on the one hand, while event B was true on the other. And among these participants, 70% of them committed the conjunction fallacy. As explained in Tentori, Bonini & Osherson (2004), exercising the conjunction rule does not require logical equivalence between *and* and ∧, it only requires that the interpretation of the *and* statement implies the corresponding ∧ statement. For example, consider the statement "he turned on the switch and the motor started" (Hertwig, Benz & Krauss, 2008p. 747). Here the use of *and* expresses a conjunction between two events, but also a temporal and causal relationship, which goes beyond the meaning of ∧. However, if the reader recognizes that both sentences "he turned on the switch" and "the motor started" must occur for the statement to be true, then the meaning assigned to *and* includes the logical operator ∧, and consequently the conjunction rule can be "properly invoked as a norm" (Tentori & Crupi, 2012, p. 124). For a full review on the misunderstanding of *and* specifically devoted to the conjunction fallacy, see Moro's review, 2009.

Pogue, Kurumada and Tenenhaus (2015) believe that conjunction fallacies are not fallacies at all, and suggest that we should view them instead as participants "behaving according to basic assumptions about the rationality of language users" (p. 2). They propose that human language does not normally consist of a series of disconnected remarks, thus when someone provides certain information (e.g., Linda is a feminist) they must have done so with a purpose. The authors agree with Hertwig and Gigerenzer (1999) that the tendency is for people to engage in goal-directed acts of communication, even with regards to the simple Linda problem, rather than treating the scenario as an abstract logical problem. If this is in fact the case, then these 'reasoning errors' are in fact grounded behaviours that lead to more successful communication.

The representativeness heuristic can only apply when "a question asks about the probability of membership of an instance in a conjunctive category and when knowledge about representative members of that category is available" (Costello & Watts, 2017, p. 305). However, the conjunction fallacy occurs frequently even when this condition does not hold. Work done by Tentori, Bonini and Osherson (2004) disputed the traditional ‘representativeness heuristic’ explanation for the conjunction fallacy, and they have shown that the conjunction fallacy occurs frequently even when there is no question about membership of an instance in a category, thus representativeness cannot explain the occurrence of the fallacy in these cases. The authors asked people to bet on the occurrence of unique future events or on the conjunction of those events. Eighty percent of their participants chose to place monetary bets on the conjunction of events even with the disclaimer: “both events must happen for you to win the money placed on this bet” (p. 468).

The fact that the conjunction fallacy is a robust, consistent and systematic part of people’s probability judgements is attested in many studies over at least the past 40 years. In some cases, these fallacies occur very frequently (e.g., rates of 80% or higher, Tversky & Kahneman, 1983), while in other cases the fallacies are much rarer (e.g., rates of 10% or less, Fisk & Pidgeon, 1996). The conjunction fallacy is an undeniable part of human probabilistic reasoning; however, it does not occur at the same rate for all conjunctions. The violation of the conjunction rule occurs very frequently with the Linda problem, after all that is what it was designed to demonstrate, but occurs significantly less frequently for other material. The lower the probability of the less probable constituent P(T), and the higher both P(F) and the conditional probability P(T&F), the more frequent the conjunction fallacy is (Fisk & Pidgeon, 1996). Indeed, these authors were able to show a decrease in the conjunction fallacy to rates as low as 10% when they manipulated both constituent events to be low in probability. Costello and Watts (2017) explain this wide range of observed fallacy rates by a model where people reason according to probability theory but are subject to random noise in the reasoning process. This model “rejects the idea that people estimate probabilities using heuristics and instead assumes that people reason according to probability theory but are subject to random noise in the reasoning process; in the model, this random noise causes the systematic biases seen in people’s probability estimates” (Costello & Watts, 2017, p. 305). In other words, probability judgements are produced by a fundamentally rational mechanism, which is perturbed in various ways through purely random noise or error.

Dual Process Theories

Since the Linda experiment, significant scholarly work has been dedicated to the problem. The phenomenon itself is robust, with experimental evidence ranging across sectors such as betting on sports games (Erceg & Galić, 2014) to investing in complex financial products (Rieger, 2012). However, the mechanisms are less understood. Traditionally, dual process theories have been used to explain why people fall prey to such reasoning fallacies. Dual process theories of reasoning differ somewhat in their details; however, they all posit that there are two distinct types of reasoning. The idea of two systems of thought is not new; it has been around for as long as academics have written about the nature of human thought. However, only in more recent years have cognitive scientists proposed that there are two distinct cognitive systems underlying thinking and reasoning with separate evolutionary histories (Evans, 2003). The animal-instinctual brain and the human-logical brain. These two systems of thought have had many labels: Heuristic and Analytic (Evans, 1984), Implicit and Explicit (Evans & Over, 1996), System 1 and System 2 (Stanovich & West, 2000), and more recently Type 1 and Type 2 (Evans, 2008). In this thesis I will use the terminology System 1 and System 2 to describe the types of thought behind human judgement and decision making according to dual process theory.

The defining feature of *System 1* processing is its autonomy (Stanovich, West & Toplak, 2011). The activation of System 1 is automatic (both processing and activation) and fast, it is not reliant on high-level control systems, thus it does not put a heavy load on central processing capacity. It is the animal brain (i.e., universal cognition shared between humans and animals), acquired through biology, exposure, and personal experience. System 1 processes are parallel and "automatic in nature: only their final product is posted in consciousness" (Evans, 2003). It tends to solve problems involuntarily and effortlessly by using prior knowledge and beliefs (Kahneman, 2011). System 1 processes are intuitive, have implicit inferences and use heuristic processing when making judgements or decisions. Unlike System 1, *System 2* is nonautonomous. It is "believed to have evolved much more recently and is thought by most theorists to be uniquely human" (Evans, 2003). It is a relatively slow, controlled and conscious processing system that is computationally expensive as it draws from cognitive resources. It solves problems using logical reasoning and its operations involve "choice and concentration" and are "heavily demanding of people's computational resources" (Kahneman, 2011, p. 21; De Neys, 2006, p. 428). It is rational, rule-based, and has explicit inferences. System 2 processes are sequential in nature and permits abstract hypothetical thought that cannot be achieved by System 1. One of the most critical functions of System 2 processing is to monitor and override System 1 processing (Stanovich, West & Toplak, 2011). See Table 1. 1 for an overview of the properties of System 1 and System 2 processes.

Table 1. 1

The properties and attributes of dual process theories of reasoning (adapted from Stanovich & West, 2000 and Evans & Stanovich, 2013)

System 1 process (intuitive)	System 2 process (reflective)
Defining features	
Does not require working memory	Requires working memory
Autonomous	Cognitive decoupling; mental simulation
Properties	
Fast	Slow
High capacity	Capacity limited
Parallel	Serial
Nonconscious	Conscious
Holistic	Analytic
Biased responses	Normative responses
Automatic	Controlled
Associative	Rule-based
Experience-based decision making	Consequential decision making
Independent of cognitive ability	Correlated with cognitive ability
Evolved early	Evolved late
Similar to animal cognition	Distinctly human
Implicit knowledge	Explicit knowledge
Basic emotions	Complex emotions
Task construal	
Contextualised	Abstract
Personalised	Depersonalised
Conversational and socialised	Asocial

Heuristics are defined as automatic processes because they occur without intention, effort or awareness, and without interfering with other concurrent cognitive processes; whereas, controlled processes are intentional, under the individual's control, effortful and entail conscious awareness (Devine & Monteith, 1999). Generally, it is accepted that System 1 processing yields fast, frugal and correct conclusions, although sometimes its reliance on heuristics and prior beliefs lead to irrational responses. When this happens System 2 is needed to suppress this initial incorrect response and replace it with a correct one. This inhibitory-control mechanism is a critical prerequisite for deliberate processing and is an idea that has been around for many decades, indeed Thurstone (1927) referred to it in his discussion on the nature of intelligence. Often people act as cognitive misers (Fiske & Taylor, 1991) by engaging in attribute substitution, which is the substitution of a difficult-to-evaluate characteristic for an easier one, even if the easier one is less accurate (Kahneman & Frederick, 2002). This model of attribute-substitution "focuses on the phenomenon whereby people often answer difficult questions by substituting an answer to an easier question without necessarily being aware of the substitution" (Böckenholt, 2012, p. 389). This generally serves people well in many situations; however, when we are evaluating important risks (e.g., for example the risk of certain activities and environments for our children) we do not want to substitute careful thought. Instead, we want to employ System 2 processing to override and block the attribute substitution of the cognitive miser. Kahneman and Frederick (2002) provide an example of how attribute substitution can lead people into irrational response patterns. In an experiment, participants assigned a higher probability to "an earthquake in California causing a flood in which more than 1000 people will drown" compared to "a flood somewhere in the United States in which more than 1000 people will drown." This is because the image of a Californian earthquake is very accessible, and its ease of accessibility affects the probability judgement.

In order to override System 1 processing, System 2 processing must be capable of two things: 1) interrupting System 1 processing and suppressing its response tendencies, and 2) replacing the erroneous response with a better one (Stanovich, West & Toplak, 2011). But where do these better responses come from? Evans (2007, 2010) suggests that they come from hypothetical reasoning and cognitive simulation, which are unique aspects of System 2 processing. Hypothetical reasoning involves creating temporary models of the world and testing our actions in that simulated world (Stanovich, West & Toplak, 2011).

At the heart of dual process theories lies a "corrective" view on sound reasoning and deliberation. In other words, correct responding is assumed to require correction of an intuitive System 1 response by slower and more deliberate System 2 processing (Kahneman, 2011; Kahneman & Frederick, 2005). This corrective view results in a somewhat indisposed characterization of System 1. It is seen as a source of error that requires supervision from the deliberate System 2. “Bluntly put, System 2 is portrayed as the good guy that cleans up the mess left behind by the fast but error prone System 1” (Bago & De Neys, in press, p. 2). To clarify, dual process theories do not suggest that intuitions are always incorrect. In fact, they are sometimes seen as both appropriate and helpful (Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996). Also, it is not claimed that deliberation will always lead to the correct answer (Evans & Stanovich, 2013; Kahneman, 2011). However, intuitive responses do have the potential to be biased and as a result must be monitored and sometimes corrected by System 2 processing. It is this corrective aspect of the model that puts System 1 processing in a negative light. However, it has not always been this way. Historically, mathematicians and physicists have conceived intuition as guiding intellect. Great minds such as Isaac Newton, Henri Poincaré and Albert Einstein placed great trust in their intuition and credited it with many of their major breakthroughs. However, this does not discount the value in deliberation. After the initial insight of the intuitive idea one must employ deliberate thought to validate and develop it further. This view differs to the corrective dual process view as here sound reasoning is not conceived as a process that overrides and corrects erroneous intuitions, but rather as a process that builds on correct insights (Bago & De Neys, in press). Following a series of experiments, the Bago and De Neys (in press) found convincing evidence for logical intuitions, or correct intuitive responding in the initial response phase. They also conclude that in addition to overriding and correcting an incorrect intuition, deliberation is often also used to verify and justify a correct intuitive insight.

Why do people commit the conjunction fallacy according to dual process theories? Different versions of dual process theories all agree that heuristic assessments originate from intuitive processing. They differ, however, in their understanding of the origin of logical assessments and people's ability to detect conflict between heuristic and logical assessments. Several propositions have been set out in past research to explain why people are prone to committing the conjunction fallacy. Some researchers believe it is due to misunderstanding the task (Politzer & Noveck, 1991; Hetwig, Benz, & Krauss, 2008). The *misunderstanding the task* account proposes that people make logical errors because the task contains linguistic ambiguities or involves concepts that are not adapted to the human mind. For example, a person might interpret the conjunction *and* as "aswell" instead of "or". In this case, when they state that the conjunction is more likely than the constituent, they are not wrong. Hertwig, Benz and Krauss (2008) suggest that when performing the classic Linda task, people interpret Pr(T) compared to Pr(F&T) as Pr(T & not-F) compared to Pr(F&T). In other words, they judge that Linda is more likely to be a bank teller and NOT a feminist than she is to be a bank teller AND a feminist. In this situation, following a vignette that describes Linda as a feminist, there can no longer be a fallacy because both alternatives involve choosing a conjunction of two events; one of which is representative (e.g., feminist) while the other is unrepresentative (e.g., not a feminist). However, evidence provided by Tentori, Bonini and Osherson (2004) disputed this account. They showed that even when people were asked to evaluate Pr(T), Pr(F&T), as well as Pr(T & not-F) they still erroneously believed that Pr(F&T) was more probable than Pr(T).

The *default-interventionist* account posits that people make errors because they are not able to detect the conflict between heuristic and logical considerations (Evans, 2008; Kahneman, 2003, 2011; Stanovich & West, 2000; Kahneman & Frederick, 2002). People are happy to trust their immediate, intuitive judgement and lack the motivation or cognitive capacity to engage in effortful deliberation. However, those who can or are willing to engage in deliberative thinking may apply the conjunction probability law to their judgement, which would enable them to make a logical and correct answer to the task (Evans, 2008; Stanovich & West, 2000; Kahneman & Frederick, 2002; Kahneman, 2003, 2011). The *parallel-competitive* account proposes that people can detect conflict but are not able to inhibit the fast heuristic answer (Sloman, 1996; Tversky & Kahneman, 1983). This is when people experience "simultaneous contradictory belief". In other words, they might suspect that there is something wrong with their judgement but will make it anyway because they lack the cognitive powers required to inhibit it (Sloman, 1996; Tversky & Kahneman, 1983; Epstein, 1994; De Neys & Glumicic, 2008).

The default-interventionist view and parallel-competitive view differ in their accounts of information processing; however, they do share the core assumption that human behaviour is determined by the interplay of two systems of processing. The default interventionist view states that deliberative processing is optional (people are either motivated to solve the problem or capable of doing so) and the parallel-competitive view states that deliberative processing is systematic (despite knowing the conjunction rule, people choose the representative answer because it is the most compelling, they are not able to inhibit the persuasive intuitive answer). The default-interventionist view assumes that initially only the heuristic system gets activated, while the analytic system monitors the output and might override the heuristic system at a later stage if a conflict is detected. Evans described this process as being serial in nature (Evans, 2008) and typically quite lax (Kahneman, 2003). The parallel-competitive view believes that people make analytic considerations right from the beginning of the judgement process and are able to detect possible conflicts through the use of heuristically cued beliefs. In other words, both systems are operating flawlessly and in parallel. Both these accounts agree that the conjunction fallacy is a fast, automatic, and persuasive answer, but that it can be overcome by employing a slower, more effortful and deliberative thinking process. In other words, heuristic considerations arise from the intuitive process, and logical considerations from the deliberative process; however, effortless, intuitive thinking can sometimes be aligned with logic and probability theory (Saxe, 1988; Villejoubert, 2009, De Neys, 2012).

Fuzzy trace theory assumes that judgements are guided by a range of preferences for types of information, and that they default to the simplest gist representation. If this fails to yield a decision, then more complex representations are used. Gists represent the essential meaning of a situation, and intuitions are believed to arise from gist memory traces (Reyna, 2012). This theory was adapted from dual process theories; however, it differs in that it distinguishes between impulsivity and intuition, which are usually combined in System 1 according to traditional dual process theories. One defining feature of Fuzzy Trace Theory is that it claims that expertise and advanced cognition both rely on intuition, it just depends on the type of representation used to process the information: gist or verbatim (Reyna, 2012). Gist and verbatim representations are activated in parallel, however people will generally rely on the simplest and least precise gist representation necessary when making decisions. Fuzzy trace theory posits that people make errors when classes overlap, when a compelling conclusion comes to mind, and when individuals fail to inhibit this pre-potent intuitive response (Reyna & Brainerd, 2008; Wolfe & Reyna, 2010). This dependence on gist makes reasoning vulnerable to processing interference from overlapping classes of events, although it also explains expert reasoning such that a person can treat certain different reasoning problems in the same way if the problems share an underlying gist (Reyna, Lloyd, & Brainerd, 2003).

The accounts mentioned above differ slightly from one another; however, they all assume that logical answers emerge form slow, deliberative, and rule-based processes whereas heuristic answers arise from faster, intuitive and associative processes. Most believe that if a reasoner relies on intuitive processes to make a judgement, that judgement will fall prey to heuristic biases. However, if a reasoner is able to suppress the prepotent heuristic response and employ deliberative reasoning, s/he will most likely arrive at the correct answer. Although, when it comes to conjunction probability judgements, the majority of the time erroneous judgements are produced as the stereotypical heuristic answer is very persuasive. This account has contributed to the widespread belief that traditional logical or probabilistic considerations play little role in our reasoning (Gigerenzer, 1996). More recently, the *logical-intuitionist account* suggests that our intuition does not always mislead us. It can sometimes reflect a sensitivity to the logic of sets which underpins the conjunction rule (Villejoubert, 2009; De Neys, 2012). In fact, there is evidence supporting the idea that fast, automatic processes reflect a sound appreciation of logic and mathematics. A study in Brazil showed that children who lived on the streets and sold candy for a living could outperform schooled children by using candies and money instead of pen and paper (Saxe, 1988). Despite having no formal schooling these candy sellers were able to perform advanced mathematical equations automatically. The children's mathematical skills emerged as they worked their trade to achieve the primary goal of economic survival. This suggests that people can acquire the laws of logical problem solving by simply existing in and interacting with the world. In other words, it is possible that people can have logical intuitions.

But what of the nature and procedure of learning complex mathematics? Could logical intuitions be likened to training effects, or automaticity, in complex mathematical problems? A study using an fMRI conducted by Delazer et al., showed that learning complex arithmetic induced a modification of cerebral activation patterns (2003, p. 84). Participants were faster and more accurate in trained problems compared to untrained problems, plus they showed a change in brain activation patterns with increasing expertise. The study also highlighted the demands on working memory in untrained problems, supporting the notion that calculation strategies rely more heavily on executive resources than automatized retrieval. Developmental research on the acquisition process of arithmetical competence in children has shown that “during the first school years, calculation processes change from deliberate and sometimes effortful step-by-step procedures to fast and efficient processing” and following “years of training, older children show increasing automaticity in retrieval of arithmetic facts” (Delazer et al., 2003, p. 77). Taken together, these studies provide support for the idea that mathematical reasoning can be automatized. However, these automatizations of knowledge may not be the same thing as logical intuitions. Logical intuitions are knowledge that is implicit in nature and activated automatically when faced with a reasoning task (De Neys, 2014). While a logical intuition is indeed a fast, automatic response, it is not simply reiterating explicitly taught information that has become automatic through practice and expertise. Exactly how one acquires this knowledge is unclear; however, it might be possible that we acquire it by simply existing in our world (e.g., Saxe, 1988). Through interacting with our environments and surroundings, we implicitly learn the fundamental laws of logic and physics that govern our world. De Neys (2012) postulates that proof of logical intuitions lies in the successful nature of conflict detection. The idea being that in order to detect conflict between intuitively cued heuristic responses and logical principles, these two types of responses need to be activated at a similar level. While Trippas, Handley, Verde and Morsanyi (2016) suggest that people have an implicit sensitivity to logical structures, which potentially occurs automatically and outside of awareness.

As the body of literature grows, there is increasing evidence that shows people can make fast and logical judgements, and they can intuitively detect the conflict between logical and heuristic considerations (De Neys, 2012, 2014, 2019; Villejoubert, 2009). Recent work on conflict detection has provided further support for logical intuitions. De Neys (2012) claimed that people have implicit knowledge of logical and probabilistic normative principles, which are automatically activated when faced with the classic reasoning problems. Thus, our intuition does not always mislead human judgement and decision making. As outlined in several studies, heuristics – or intuitive thinking – are helpful and often lead people to make a correct judgement (see Saxe, 1988; De Neys, 2012; Villejoubert, 2009). However, it is important to note that despite being sensitive to the fact that the heuristic answer conflicts with logical norms, the majority of people continue to give the incorrect answer when performing reasoning problems.

Evidence for logical intuitions. In 2009, Villejoubert designed an experiment to better understand the cognitive underpinnings of people's intuitions concerning the conjunction rule of probabilities. The study was intended to test the competing predictions of the lax-monitoring view (Kahneman, 2003; Kahneman & Frederick, 2002) and the conflict-monitoring view (De Neys & Glumicic, 2008) of cognitive functioning. It aimed to answer the question of whether intuitions were solely driven by heuristic processing (lax-monitoring view), or whether they resulted from a combination of heuristic and controlled processing (conflict-monitoring view). Using a novel methodology adapted from Evans and Curtis-Holmes (2005) that disentangled the two systems of thought she was able to assess the impact that time pressure (e.g., a deadline) had on the logicality of peoples' conjunction judgements. This methodology incorporated the use of both congruent statements (i.e., no conflict between the two systems of thought) and incongruent statements (i.e., conflict between the two systems of thought), allowing for the detection of conflict sensitivity. It also disentangled the systems of thought allowing the researcher to observe whether the participant was answering in accordance with representativeness (i.e., the heuristic stereotype), or logic (i.e., the rule of conjunction probability). Based on the theoretical predictions of dual-process theories, the author expected to find that people who made their judgements under time pressure would show an increased level of heuristic processing and a reduced level of logical responding. The study used a between-subjects design, and in the time pressured condition participants were allocated 12 seconds to read the statement and make an answer. The other condition had unlimited time. Villejoubert found that participants were more likely to accept representative statements (i.e., in line with the stereotype), but they were also significantly less likely to accept illogical ones. Moreover, they consistently took longer before accepting a statement when intuitive and deliberative considerations were in conflict, even when they were pressured to provide an answer in less than 12s. The increase in acceptance time under conflicting situations suggested that people engaged effortlessly in deliberative processing (i.e., that people instinctively wanted to think hard about their response when they suspected two contradictory answers). Indeed, participants appeared to be aware of the conflict and were not blindly relying on the heuristic intuitive response to guide their answers (i.e., in which case if they were there would have been no difference in time latencies between no-conflict and conflict trials). Additionally, the fact that logical considerations were still effective under time pressure suggested that the conflict between intuitive and deliberative considerations was readily detected. In other words, "intuitions regarding the conjunction rule of probabilities are not solely influenced by heuristic processing and that people think longest when there is a conflict between the heuristic-based response and the logic-based response" (Villejoubert, 2009, p. 2985). The author also found that contrary to common two-systems assumptions, representative assessments were not fast and automatic, while logical responses were not slow and deliberate. Surprisingly, when participants were given unlimited time to think about their answers, they were most influenced by the representativeness heuristic, but when put under time pressure to respond quickly, participants were in fact providing more logical responses in line with the statistical rule of conjunction probability. Villejoubert suggests that one possible account of these results could be that intuitive thinking might in fact point to the logical answer first; however, when this answer conflicts with representativeness, people then engage in deliberative thinking to resolve the conflict. This counterintuitive hypothesis goes against all current accounts of dual process theories (as reviewed above except for the logical intuitions account) since they all stipulate that when the slow deliberative system is not given enough time to process information, thinking will be more heuristic and less logical. In contrast, these findings are in line with the conflict-monitoring view (De Neys & Glumicic, 2008) and seem to suggest that fast thinking can be logical while slower, more effortful thinking can be biased by non-logical heuristics. It is important to note here however, that a time limit of 12 seconds is very generous for the limited time condition. In the allocated 12 seconds timeframe, it is possible that participants could have read the statement and still had enough time remaining to deliberate their answers. Future studies should consider severely constricting the timeframe of the limited time condition.

An important and interesting study conducted by Riis and Schwarz (2003) showed that increased heuristic processing led to fewer conjunction fallacies, while deliberate processing led to higher fallacy rates. The authors conducted two experiments using two different conjunction problems. The first was the Linda problem (Kahneman & Tversky, 1983) and the second was the James problem (adapted from Mellers, Hertwig & Kahneman, 2001). A similar type of problem to the Linda task with a personality vignette followed by ranking two statements: one single event and one conjunction event. Participants either used the approach action (e.g., pulling something closer and flexing the arm), the avoid action (e.g., pushing something away and extending the arm), or no action, while they completed the conjunction problems. The approach action has been shown to enhance heuristic processing, while the avoid action has been shown to enhance analytic processing (Friedman & Förster, 2000). The authors found something quite unexpected; participants who engaged in avoidance motor actions were significantly more likely to commit the conjunction fallacy than participants who engaged in approach motor actions. In other words, enhanced analytic processing led to higher rates of the conjunction fallacy than enhanced heuristic processing. The authors suggest that "detail-oriented processing can lead participants to pay too much attention to irrelevant details in a given narrative, which can undermine consistency with normative principles of judgement" (p. 247). Another interpretation of these findings could be that people have an implicit understanding of the conjunction law and this is obvious when they respond intuitively to the problems (i.e., logical intuitions); however, when they deliberate their answers, that is when the bias (e.g., the representativeness heuristic) corrupts their judgements. Indeed, the authors noted that deliberation was "detrimental to performance on problems of the Linda type" (p. 259).

De Neys had been doing extensive research on conflict sensitivity (De Neys, Vartanian, & Goel, 2008; De Neys & Glumicic, 2008; De Neys, 2009; Franssens & De Neys, 2009; De Neys, Moyens, & Vansteenwegen, 2010; De Neys, Cromheeke, & Osman, 2011), and in 2012 he published an article that presented people were able to implicitly detect when their heuristic response conflicted with traditional normative considerations. Following on from this conflict sensitivity he proposed that logical and probabilistic knowledge is intuitive and is activated automatically when people engage in a reasoning task. "The conflict detection studies established that despite the well documented failure to give the correct answer on the classic problems, people do not simply disregard the traditional normative implication of their judgments; rather, they are sensitive to the fact that their heuristic answer conflicts with it" (De Neys, 2012, p. 30). The author proposes that the nature of the normative information needed to detect conflict is intuitive: it is activated automatically, and it is implicit in nature. De Neys highlights the fact that dual process models lack a way to detect whether deliberate thinking is required without having to engage in deliberate thinking, and in an attempt to solve this conceptual puzzle he presents the *Logical Intuition Model*. This model proposes that intuitive-heuristic (e.g., responses are based on mere semantic and stereotypical associations) and intuitive-logical (e.g., responses are based on the activation of traditional logical and probabilistic normative principles) processing are both cued initially, if there is conflict detected between these two processing systems, then deliberate processing is engaged. If the two intuitive responses are consistent with each other, then that response will be chosen and there will be no need for further deliberation. Whereas the employment of intuitive-heuristic and intuitive-logical processing is mandatory from the start, the employment of deliberation is optional. However, the fact that deliberate processes are called upon does not imply that the correct response will be chosen. De Neys claims that "despite the erroneous answer, people have implicit knowledge of the logical and probabilistic normative principles that are evoked in the classic problems and automatically activate this knowledge when faced with the reasoning problem" (2012, p. 36).

Recent work by Bago and De Neys (in press) has shown convincing evidence that logical intuitions, or correct "intuitive" responding in the initial response phase, exist. They adopted the two-response paradigm designed by Thompson, Prowse, Turner and Pennycook (2011) to gain direct behavioural insight into the timing of intuitive and deliberative response generation. This paradigm requires participants to answer a reasoning problem as quickly as possible with the first, intuitive response that comes to mind. Following this, they are presented with the same problem again, but given as much time as they want to think about their final answer. The authors paired this two-response paradigm with items directly modelled after the bat-and-ball problem and conducted a series of experiments to test the role that intuition plays in the bat-and-ball problem. The bat-and-ball problem is a classic conflict detection problem, and one of the three original problems found in Frederick's (2005) Cognitive Reflection Task (see p. 55 for an example). It was designed to entice a potent intuitive response and requires mental capacity in order to inhibit this intuitive response and replace it with a deliberative one. In their first experiment, they employed stringent procedures to ensure that the first response given was intuitive in nature. These procedures included extreme time pressure in the form of a 4 second deadline and overburdening of cognitive resources using a secondary load task (in this instance the dot memorization task was used, Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). The authors found that the majority of participants fail to solve the bat-and-ball problem. However, in the cases where people did manage to give a correct final response following deliberation they had often already selected this answer as their first (e.g., 67.2%), intuitive response. This suggests that correct responders do not need to correct their intuition as their intuition is often already correct. This high non-corrective rate argues against the dual process theories that assume System 2 processing plays a corrective role. In fact, these findings suggest that in these cases, deliberation acts to verify correct intuitive insights. These results were replicated across a series of five experiments. In their final two experiments they explored the participants justifications of their answers (in both response phases) by asking the following questions: “Could you please try to explain why you selected this answer? Can you briefly justify why you believe it is correct? Please type down your justification below.” The combination of asking participants to generate their own response and then justify it boosted deliberation, which resulted in the lowest recorded corrective rate of all the experiments (e.g., 30%). Additionally, the justification responses clearly showed that people can estimate the correct answer intuitively, but do not know how they did it. However, after deliberation they provided clear justifications. The authors suggest that we need to upgrade our views of both Systems of thought. System 1 is able to generate correct logical intuitions, and as such System 2 deliberation cannot exist exclusively as a correction process. Instead the authors suggest that in these situations System 2 is crucial to produce cognitive transparency, which is important for many reasons, such as developing ideas further and communication. The authors stress that this interpretation does not discount System 2 deliberation, as they found that in a few instances participants were able to correct their incorrect first response after deliberation. Also, for the majority of the time participants were not able to arrive at the correct answer in either response phases. “the vast majority of reasoners gives the faulty intuitive “10 cents” response both at the initial response stage and after deliberation. Hence, not everyone will generate a correct (intuitive) response. For most reasoners, the incorrect intuition will dominate” (Bago & De Neys, in press, p.21)

In a recent study by Trippas, Handley, Verde and Morsanyi (2016) it was shown that people have an implicit sensitivity to logic, which occurs potentially automatically and outside of awareness. Across three experiments, the authors asked participants to rate sentences for how much they liked them, and how physically bright they found the sentences to be. Logically valid sentences were judged more likeable and brighter than other sentences. Although the sentences made no reference to logical validity, participants liked the statements more when they formed part of a logically valid argument, suggesting that people are sensitive to logical structure at an intuitive level. Participants also judged logically valid statements as being physically brighter than invalid statements. The authors state that “reasoning does not seem to be exclusively deliberative but instead possesses some qualities typically associated with intuitive, heuristic processes” (p. 1454).

Conflict detection

In order for people to detect that their initial response might be biased there has to be some type of conflict-detection mechanism that alerts us to this error so that we can employ System 2 processing to suppress and override the intuitive response with a deliberate one. Conflict detection studies have focused on people’s processing of a variety of classical problems found within the judgement and decision-making literature. Some of these include base rate neglect, conjunction fallacy, belief bias syllogisms, and ratio bias tasks. Giving the correct answer to these tasks involves simply applying some basic logical or probabilistic principles. However, the problems are constructed in such a way that they cue a tempting heuristic response, that is in direct conflict with these principles.

Take the bat-and-ball problem as an example. Similar to the Linda problem, the bat-and-ball problem is designed to elicit a salient intuitive, but biased response. The classic version of the problem is as follows:

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

The majority of people answer this problem incorrectly by assuming that the answer is ‘10 cents’. If the ball costs $0.10, then the bat should cost $1.10 (because as specified in the problem, the bat costs $1.00 more than the ball), which brings the total amount to $1.20. The correct answer is that the ball costs $0.05, and the bat $1.05, which accumulates in a total amount of $1.10. Tremoliere and De Neys (2014) suggest that people use attribute substitution when attempting to solve this problem, which results in the incorrect answer. People substitute the ‘more than’ relational statement with an easier absolute statement, thus the statement no longer reads ‘The bat costs $1.00 *more than* the ball’, but instead is read as ‘The bat costs $1.00’.

The conflict detection problems are trying to establish whether reasoners are aware that their biased response is questionable, thus they are designed in a way to contrast people’s processing of the traditional problem (e.g., contains conflict) with a newly constructed control/congruent version (e.g., does not contain conflict; see Figure 1. 3 for an example of both conflict and congruent versions of the Linda problem). By removing the conflict from the task, one enables the cued heuristic response to align with the response cued by logical or probabilistic principles. This makes it very easy to respond as both heuristic and logical considerations point towards the same answer. Conflict detection studies examine whether people process the conflict and no conflict versions differently. The idea behind this being that if people are simply cognitive misers and are not sensitive to the conflict underlying these considerations, then the two versions of the task (e.g., conflict present vs. conflict absent) should be processed in the same way.

A. Control "No conflict" version	B. Classic "Conflict" version
Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations.	Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations.
Which one of the following statements is most likely? a. Linda is active in the feminist movement. b. Linda is a bank teller and is active in the feminist movement.	Which one of the following statements is most likely? a. Linda is a bank teller. b. Linda is a bank teller and is active in the feminist movement.

Figure 1. 3. Illustration of a control "no conflict" and classic "conflict" version of the Linda problem. *Note*. The left panel (A) shows the newly constructed control version of the task, in which the cued heuristic response is consistent with the logical response. The right panel (B) shows the classic version of the task, in which the cued heuristic response is at odds with the correct logical response (i.e., the response considered correct according to standard logic or probability theory principles)

Tremoliere and De Neys (2014) recently investigated the question ‘when are heuristics beneficial’ in a study that used an adapted version of the bat-and-ball problem. They presented two experiments as support for their hypothesis that prior beliefs (i.e., intuitions) provide support for solving the bat-and-ball problem. By manipulating the objects and the values of these objects they were able to present congruent or incongruent versions of the problem to their participants. An example of an incongruent problem was as follows: “A Ferrari and a Ford together cost $190,000. The Ferrari costs $100,000 more than the Ford. How much does the Ford cost?” Consistent with the structure of the original problem, the cued heuristic response for the value of the Ford is ‘$90,000’; however, this amount is unbelievable for such an ordinary car. The authors predicted that the conflict between the value of the car and people’s prior knowledge would decrease the appeal of the substituted response, and thereby help people to reason better. Their results supported their predictions, showing that participants performed significantly better on solving the incongruent compared to the congruent problem (which read: “A Ferrari and a Ford together cost $190,000. The Ford costs $100,000 more than the Ferrari. How much does the Ferrari cost?”) In the congruent version, prior belief of a Ferrari being expensive would fit the cued response; however, there was a limitation in this study because prior knowledge would make it unbelievable to accept that a Ford could cost $100,000 more than a Ferrari. Thus, the authors attempted to consolidate this limitation with a second experiment that substituted the Ford with a Rolls-Royce. They were able to replicate their results from the first experiment showing that participants performed better on the incongruent (unbelievable, intuitively cued response conflicted with background knowledge) version than the congruent (believable, intuitively cued response matched with background knowledge) version of the task. However, this effect was small resulting in only a 10% accuracy boost. They attempt to explain their findings in two ways. The first is that people are readily able to detect conflict.

They typically follow their intuition when the response is congruent with their beliefs but will switch to a more deliberate thinking style when the response conflicts with their beliefs. The second explanation originates from the fluency framework, which suggests that fluency decreases when prior beliefs conflict with the intuitive response, resulting in more effortful thinking.

Feelings of rightness/wrongness. Consider again the classic Linda problem in Figure 1. 3. The vast majority of people pick the heuristic response the first time they encounter these tasks. However, this does not mean they were fully convinced that their answer was correct. The persuasive heuristic response might have prevailed; however, did respondents feel there was something wrong with their answer? Did they "feel that there was something tricky about the problem" (De Neys, 2012). The point of these feelings of rightness, or conflict detection, is to enable a "switch" from System 1 to System 2 thinking, to allow ourselves to override the intuitive response and replace it with a deliberate one. Koriat (1993) labelled the idea of switching between systems of thought as the "feeling of rightness" (FOR). The idea being that our intuitive System 1 judgements are always accompanied by an affective FOR response. This response is supposed to arise from the fluency with which the initial answer is produced. In other words, fluent answers give rise to a strong FOR, while dysfluent answers result in negative affect and lower FOR. A low FOR triggers the need to employ System 2 thinking.

The fundamental question that conflict detection studies have been trying to answer is: are people sensitive to the intrinsic conflict between the cued heuristic response and the basic logical or probabilistic principles that are evoked in these types of problems, regardless of the eventual response they produce? Recent work on conflict sensitivity, even during biased reasoning, suggests that people are able to sense when something is not quite right with their response, which in turn causes them to question their response (e.g., Bonner &

Newell, 2010; De Neys, Cromheeke, & Osman, 2011; De Neys, Moyens, & Vansteenwegen, 2010). The following studies all showed that people are sensitive to conflict and might be aware of a particular alternative answer. For example, giving an unwarranted heuristic response to a task resulted in autonomic arousal in skin conductance (e.g., De Neys et al., 2010), longer response times (e.g., Bonner & Newell, 2010; Villejoubert, 2009; De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2012), and increased attention and inspection of logically critical problem parts (Ball, Philips, Wade, & Qualyle, 2006). Additionally, the following studies show a more general lack of confidence, with participants expressing less subjective confidence after solving the classic conflict version of a problem compared to the congruent version (e.g., De Neys & Glumicic, 2008; De Neys et al., 2011; De Neys, Comheeke, & Osman, 2011; De Neys & Feremans, 2013). Taken together, these results suggest that both biased and unbiased reasoners are readily able to detect conflict and can sense that their heuristic response is questionable. Failure to provide an unbiased answer is not due to a lack of conflict detection, as both biased and unbiased reasoners showed evidence of conflict sensitivity.

The origins of conflict detection. Recently in the body of literature surrounding conflict detection a debate has arisen regarding how cognizant conflict detection actually is. To be more specific, empirical studies present substantial support for the presence of conflict detection. They conclude that even biased reasoners are able to detect that their incorrect response is questionable (Bago et al., 2019; Villejoubert, 2009; De Neys & Glumicic, 2008). When people give a heuristic response that conflicts with the laws of logic, they are able to pick up on this at some level, and this is where the problem lies. Researchers are aware of its existence; however, the precise nature of conflict detection is unknown. A burning question remains: where does this doubt come from? Bago and colleagues (2019) suggest two possibilities. The first is that "people have a highly specific error or conflict signal" (p. 215).

Take the bat-and-ball problem as an example. With regards to conflict detection, people might have computed both the “10 cents” and the “5 cents” responses, thus they experience a conflict between the correct and incorrect response. Biased reasoners would choose the heuristic response of “10 cents”; however, they would also be aware that the alternative answer is “5 cents”. The second suggestion is that “the conflict signal might be non-specific” (p. 215). In other words, people might detect that the heuristic answer of “10 cents” is incorrect, but they would have no further clue as to what the correct response is.

A few recent studies have attempted to illuminate where conflict detection comes from. Travers and colleagues (2016) used a mouse-tracking methodology to try and pin point where the conflict was coming from. They used an adapted version of the extended CRT and designed both conflict and no-conflict trials. A choice of four different responses were presented in the corners of the screen. Conflict trials contained both the heuristic response and the logical response, plus two addition foil responses. No-conflict trials contained the logico-heuristic response plus three foil answers. The researchers found that while correct responders showed attraction towards the incorrect “10 cents” option, incorrect responders showed no attraction towards the correct “5 cents” option. These findings either mean that contrary to other conflict detection studies, biased responders do not display error sensitivity. Alternatively, it might indicate that biased responders conflict detection is non-specific in nature. To be more specific, reasoners might detect that the “10 cents” response is erroneous; however, do not know that the “5 cents” response is correct, hence, showing no particular attraction towards it.

Bago and colleagues (2019) attempt to further clarify this issue in their study by introducing a second guess paradigm. First, they asked participants to answer an adaptation of the classic bat-and-ball problem. Following this, they asked them to make a second guess, which they could choose from a list of options (e.g., “1 cent”, “5 cents”, or “15 cents”). They purposefully left the intuitively induced response off the list of second guess options as a means to reflect the specificity of their conflict detection. The researchers found over a series of three experiments, that the "average biased reasoner has a medium-specific error signal" (p. 225). What this means is that biased participants were not aware of the correct response and gave second guess answers lower than but as close to the heuristic answer "10 cents" as possible. Although they did not know the correct response, they were aware that it should at least be smaller than the salient heuristic answer they succumbed to. However, it should be mentioned that this evidence is weak and does not allow for the possibility of participants guessing. The task was a forced answer task from three different choices, two of which were lower, or smaller than, than the heuristic answer. Thus, giving participants a 2 out of 3 chance (i.e., 66%) of guessing a lower number than the salient heuristic answer.

Implicit vs. explicit detection. People are readily able to detect the conflict underlying logical and heuristic considerations (De Neys & Glumicic, 2008; Villejoubert, 2009; Franssens & De Neys, 2009), but to what extent are these realizations conscious? De Neys and Glumicic (2008) found that explicit verbalization did not show any evidence of actively experienced conflict; however, analysis of processing levels from a recall test showed that they were experiencing this conflict implicitly (i.e., conflict detection was accompanied by deeper processing of the base rate information, which resulted in higher recall). In other words, "people are not verbalizing the conflict they are detecting" (p. 1274). However, it is important to note that lack of verbalization does not simply imply that the detection process is unconscious. The role of verbalization as a key prerequisite for conscious, explicit processing is debatable. Moors and de Houwer (2006) caution against using dichotomous implicit-explicit labels, and instead favour a more gradual approach. With this in mind, De Neys and Glumicic (2008) suggest that "the conflict experience might be less explicit than traditionally assumed but any stronger claims should be avoided" (p. 1275).

Conflict detection and dual process theories. Thanks to the advancement and accessibility of technology inspiring new methodologies, conflict detection research is thriving, and people are learning a lot more about dual process theories than ever before. For example, the fact that conflict monitoring is successful even when the biased response is given suggests that the dominance of the heuristic response is not due to a lax monitoring process (Bago et al., 2019; Travers, Rolison, & Feeney, 2016; Villejoubert, 2009; De Neys & Glumicic, 2008). This implies that people are not simply cognitive misers, instead we have acquired the relevant normative principles needed to detect the conflict underlying heuristic and logical considerations, and we are able to retrieve them. If people were not taking analytic considerations into account, then they would simply not detect that there is conflict, which does not happen. Instead, the dominance of heuristic responses should be accredited to an inhibition failure.

It is apparent from the literature that the nature of the relation between the two systems is unclear (Evans, 2007, 2009). Is it a serial or a parallel relationship? The parallel activation account (e.g., Epstein, 1994; Sloman, 1996) states that both systems are activated from the start and they simultaneously compute a solution. Hence, if their answers differ they are alerted to the conflict between intuitive and logical considerations. The serial view differs in that a reasoner initially relies on only the intuitive system. The deliberate system is only employed when the intuitively cued response conflicts with the output of the deliberate system. There is; however, a fundamental flaw in the conception of the serial view. How is the reasoner able to detect a conflict between the output of the intuitive and deliberate system if the deliberate system is not engaged yet? The simultaneous activation of the two systems in the parallel view overcomes this problem; however, it also faces a problem as the deliberate system is "blindly" engaged from the beginning (De Neys, 2012). Thus, essentially throwing away any benefit of having an intuitive system. When the two systems of thought do not conflict with each other, it is perfectly acceptable to rely on the intuitive route. Engaging in demanding deliberations in this case is redundant and would be a waste of cognitive resources (De Neys & Glumicic, 2008). De Neys (2012, 2013) has postulated that in order to detect conflict between intuitively cued heuristic intuitions and logical principles, the logical principles need to be activated at some level. Conflict detection has an effortless nature (Franssens & De Neys, 2009), thus it seems to be a System 1 process. However, system 1 is widely accepted as being intuitive. So, how are people able to detect the conflict between heuristic and logical considerations when only System 1 (i.e., intuition) is activated initially? Activating logical considerations from the beginning would be a waste of precious cognitive resources, thus any realistic dual process model needs a way to detect whether System 2 thinking is required without having to engage in System 2 thinking (Evans, 2009; Thompson, Turner, & Pennycook, 2011).

De Neys (2011), suggests that the cuing of an intuitive logical response is the answer to this conceptual puzzle. If the intuitive System 1 cues both a heuristic and logical response, then any potential conflict can be detected without prior engagement of System 2. This involves a parallel activation, not of the two systems as in the parallel-competitive account, but rather a parallel activation of two different types of intuitive responses: a heuristic-intuitive response based on semantic and stereotypical associations, and a logical-intuitive response based on the activation of traditional logical and probabilistic norms (De Neys, 2012). In the case where no conflict exists (i.e., congruency between the two intuitive responses) people will simply select the cued response without further deliberation. However, any detected conflict would result in the need to engage the deliberative system. Importantly, this does not necessarily mean that the correct logical answer will be chosen, but this account does provide us with a clear switch rule to understand how conflict might be detected without permanently activating the deliberate system. De Neys (2012) does clarify that the logical intuition account differs from the basic automatization account, because all reasoners have access to these logical intuitions, not just experts (e.g., a professional statistician might be able to solve logical reasoning problems in an entirely intuitive manner after years of extensive training). Additionally, De Neys clarifies that the intuitively cued response might still be chosen despite detecting the conflict because “the heuristic response might be more strongly activated, salient, or appealing than the logical response” (2012, p. 35). There is no guarantee that the two intuitive responses have the same strength or status. If conflict is detected and deliberation is henceforth activated, then the final selection of the logical response will still require a demanding inhibition of the heuristic response.

The present research

Dual process models do not necessitate that correct responding requires System 2 processes. Indeed, the correct response can come from System 1 processing when there is no conflict between the two systems of thought, but also when the task is elementary enough. For example, “Modus Ponens is arguably automatic, and, as long as the conditional major premise is encoded in working memory, it can be triggered unconsciously by a minor premise that is below the threshold of detection” (Reverberi, Pischedda, Burigo, & Cherubini, 2012, p.252). Modus Ponens is an intuitively straightforward rule of inference. It can be summarised as “A implies B and A is asserted to be true, therefore B must be true. In this case, the processing needed to arrive at the correct conclusion is so basic that it has been automatized and assimilated as a System 1 response. If there is no erroneous response coming from System 1, then there is obviously no need for System 2 to correct it. Modus Tollens is a related rule of inference to Modus Ponens. Summarised as “A implies B, but B is not true, therefore A is also not true”. For example, “If the watch-dog detects an intruder, the watch-dog will bark. The watch-dog did not bark. Therefore, no intruder was detected by the watch-dog”. Following this general idea, some might suggest that correct intuitive responses are attributed to the easiness, or non-representative nature, of the task (Pennycook, Fugelsang, & Koehler, 2012; Travers, Rolison, & Feeney, 2016). However, I believe that the prototypical Linda problem, and various adaptions thereof, is demanding enough to warrant deliberation. Particularly when there is a high contrast between the stereotypes or prototypes used in the task. For example, if *Pr*(T) is highly unrepresentative of a chosen category, but *Pr*(F) is highly representative of that category. When testing for correct "intuitive" responding, it is important to ensure that rigid measures are put in place to ensure that the first response is intuitive in nature. If simply instructed to respond intuitively, but not forced to, participants might fail to respect the instructions resulting in a correct first response due to System 2 processing. For example, in a study conducted by Villejoubert (2009), participants in the time restricted condition were instructed to respond within 12 seconds. This time limit was prescribed as it was less than half the average time taken under unlimited time in a pilot study; however, reviewers have suggested this is too long a period of time to warrant intuitive responses. Clearly one needs to make sure, through the use of stringent measures, that only System 1 is engaged in the initial response phase. It is widely accepted that System 2 processing is time and resource demanding (Kahneman, 2011; Kahneman & Frederick, 2005). "Under conditions of time pressure and memory load people are less able to recruit effortful analytical mechanisms in order to correct the prepotent heuristic response, supporting the contention that the incorrect response is based on a less thorough analysis and is delivered without much effort" (Crisp & Feeney, 2009, p.2323). Thus, in the present series of experiments, I attempt to force the intuitive response through discarding System 2 as much as possible by depriving participants of time and cognitive resources. By having participants reason under conditions that minimise deliberation in the first response, I attempt to identify the presumed intuitive response that precedes deliberation.

Can we intuit probabilities? De Neys (2012) as well as Handley and Trippas (2015) offer their explanations for believing that reasoning, an accepted System 2 process, may in fact be partly intuitive. They showed that when people make biased or erroneous judgements (i.e., they fail to make normatively correct judgements in line with logical considerations), they often reveal a sensitivity to the normatively correct knowledge through implicit measures; including response latencies, response confidence ratings and gaze durations. However, when people are in fact able to make normatively correct judgements, they often neglect to mention critical aspects of the normative solution in their verbal protocols (e.g., referring to the base rate information when solving a base rate task). These findings suggest that procedures relevant to reasoning might in fact be activated implicitly and outside of awareness – a characteristic of intuition rather than deliberation. Trippas, Handley, Verde, and Morsanyi (2016) also provided evidence that suggests people have an implicit sensitivity to logical norms. Participants were not asked to reason, instead they were asked to rate how much they liked a sentence, or how physically bright they found a sentence. Results showed that people rated sentences with inherent logical structure as more likeable and brighter than sentences that were logically invalid. The authors surmised that although reasoning is undeniably a System 2 process and resource-intensive, it also "possesses some qualities typically associated with intuitive, heuristic processes" (p. 1454).

De Neys (2012) supports this idea with his "logical intuitions" model, whereby people appear to be intuitively sensitive to the conflict between heuristic and deliberate considerations. For example, stereotypes and formal probabilistic norms. This model states that people have an intuitive sensitivity to all formal norms; however, they are often unable to articulate this conflict. The analytic-heuristic conflict is always detected, but this does not always result in correct responses being made due to inhibition failure. A few reasons for this may be that people lack the motivation or cognitive resources to complete the demanding inhibition process, or that there is no follow up of deliberate System 2 thinking (i.e., justification of the initial logical intuition) after conflict detection (De Neys & Bonnefon, 2013). Despite much evidence supporting the idea that people are able to detect the conflict between their deliberate and heuristic considerations (Stupple & Ball, 2008; De Neys, Moyens & Vansteenwegen, 2010; De Neys, 2012; Villejoubert, 2009; De Neys, Vartanian, & Goel, 2008; Franssens & De Neys, 2009; De Neys & Glumicic, 2008; Mevel et al., 2014), as well as the fact that people might have logical intuitions (Saxe, 1988; De Neys, 2012; Villejoubert, 2009), there remains a large gap in our knowledge as to whether people are genuinely answering in accordance with their logical intuitions when they make these judgements, or using a different strategy. One cannot disregard the potential explanation for these "logical" responses as resulting from logical reasoning (i.e., System 2 deliberation), from a rejection of the heuristic answer and simply choosing the alternative option (e.g., in a forced answer task containing two possible answers), or from participants simply giving up and guessing.

Can we deliberate intuitions? Recently, there has been some evidence provided for the notion of effortful heuristics, or heuristic reflections. Handley, Newstead, and Trippas (2011) used a novel methodology by asking participants to judge both the validity as well as the believability of an argument. They found that belief judgements (i.e., intuitive judgements) took longer to make and were also more subject to error than logically valid judgements. The authors concluded that the logical consideration (e.g., the validity) was available early, but needed to be inhibited in order to explicitly assess the believability of the argument. A possible interpretation of these findings is that "the logical inference is available automatically and by default and intervention is required to successfully give a belief judgement" (Howarth, Handley & Walsh, 2018, p. 5). In a subsequent study conducted by Howarth, Handley and Walsh (2016) using the same novel methodology as Handley et al.

(2011) plus a secondary task that burdened working memory resources, the results mentioned above were replicated. Belief-based judgements produced lower rates of accuracy overall (e.g., compared to validity judgements) and were more greatly influenced by conflict being present. The secondary task (i.e., random number generation) reduced accuracy across all judgements, however had its greatest impact on logic-based judgements. The authors interpreted these findings as resembling conflict between two System 2 processes. Indeed, they proposed that logical responses are available at an early System 1 level, however require System 2 processing (e.g., deliberation) to explicitly extract the underlying structure of an argument required for answering in line with logical considerations. However, belief-based responses are available later and depend on the inhibition of a competing intuitive-logical response. The authors propose that this evidence is better suited to a parallel processing model compared to a default interventionist model. A few years later, the same authors conducted another series of two experiments to investigate the effortful nature of belief-based (i.e., heuristic) judgements (Howarth, Handley & Walsh, 2018). They increased the inhibitory demands of the task by adding a third instructional condition (e.g., the Stroop Test in E1 and a memory updating task in E2) and found that this in turn increased the difficulty of withholding a logical response, therefore reducing accuracy on the belief judgement. Logic judgements appeared to impact on belief judgements more when inhibitory demands were increased by the inclusion of a colour naming condition. This led the authors to surmise that people use a parallel processing model to make judgement and decisions. They believe that simple logical considerations are available early and require inhibiting to produce a belief-based (i.e., intuitive or heuristic) response. Thus, there are two routes to providing a logical response: an intuitive route and a deliberative route. The first is a System 1 process completing first and creating an intuitive cue based on the logical structure of the argument, possibly accompanied by a feeling of rightness (Thompson, Prowse, Turner & Pennycook, 2011). The second is a System 2 process activated when given specific instruction to reason logically, which runs parallel to the route required for an intuitive/heuristic response and explains the impact of a secondary task on logic judgements. With simple logical arguments, as used by Howarth et al. (2018), an intuitive logical response is available early and must be inhibited in order to explicitly generate a response based on belief, or intuition. Their findings showed that withholding an available logical response requires effort and increasing cognitive demands significantly reduced people's ability to do so. However, on more challenging tasks a belief-based, or intuitive, response may come first, thus impacting logic-based reasoning.

The current research is focussed around the question of whether people are able to intuitively solve conjunction probability judgements. From simply existing in the world, we learn different lessons and rules. For example, Saxe (1988) found that street children who sold candy for a living were able to outperform school children on advanced mathematical equations, simply because they had learned to do so from existing in their environment. These children had received no formal schooling. Could this also be the case for people with the statistical law of conjunction probability? Indeed, the logical intuitions model (De Neys, 2012) suggests that people have an understanding of the formal laws underlying most classic reasoning and decision-making tasks (e.g., base-rate neglect, conjunction fallacy, ratio-bias, belief-bias, etc). In 2009, Villejoubert ran an experiment that resulted in some surprising data. The results showed that people took longer to make a judgement when heuristic and logical considerations were in conflict, even when they were put under time pressure. This suggests that people do readily detect the conflict between heuristic and logical assessments. Additionally, time pressure appeared to increase the rate of logical judgements under conflict. One possible interpretation of these results would be that intuitive thinking might in fact point to the logical answer first (i.e., logical intuition); however, if deliberation is employed the heuristic bias then corrupts our thinking (i.e., heuristic reflection). This idea goes against all traditional accounts of dual process theories, since they all stipulate that when the slow deliberative system is not given enough time to process information, thinking will be more heuristic and less logical. In contrast, Villejoubert's (2009) findings suggest that quick thinking can be logical while effortful thinking can be biased by non-logical heuristics.

Hypotheses. The present research proposes to build on Villejoubert's (2009) findings, and methodology, to explore more systematically the circumstances under which accurate logical considerations might be fast and intuitive in judgements of conjunction probability. This research has two objectives: (1) To determine how individual differences impact on the logicality of individuals' probability judgements, and (2) to test original and theoretically-driven ways to improve individuals' probability judgements.

The first objective will be addressed by testing two hypotheses: (1a) Individuals who tend to engage more naturally in effortful thinking are also more prone to produce illogical probability judgements if they have time to deliberate because their deliberation is biased by non-logical heuristics, and (1b) in situations of conflict between representativeness and logicality, individuals who provide logical answers will show different patterns of eye-movement disruption than individuals who provide heuristic answers.

The second objective will be addressed by testing two additional hypotheses: (2a) Decreasing individuals' capacity for effortful thinking will also have a positive impact on probability judgments because it will impede the inhibition of the rapid, logical response by conscious effortful deliberation, and (2b) the presence of conflict and severe time pressure will result in individuals' displaying more logical base-rate responses than when there is no conflict or no time pressure, because deliberation is biased by non-logical heuristics. This final experiment seeks to replicate the findings from Hypothesis 1a but using a repeated measure design and investigate whether the results are transferable across different probability tasks.

Chapter 2: Methodology

In the classic conjunction fallacy experiments (e.g., Tversky & Kahneman, 1983), most participants commit the conjunction fallacy when they rate an atypical single event (e.g., “Linda is a bank teller” (T)) as more likely than a conjunction of the atypical single event and a representative one (e.g., “Linda is a Feminist and a bank teller (F&T)). Yet, they also correctly believe that a representative single event (e.g., “Linda is a Feminist” (F)) is more likely to occur than the above-mentioned conjunction. Researchers have assumed that participants who ranked (F) as more probable than both (T) and (F&T) answered logically because they chose the single event over the conjunction. However, it is impossible to disentangle logic from intuition using this design. The rating of (F) as more likely than (F&T) could stem from either logical considerations (since a single event is more likely than the conjunction of two events) or intuitive considerations (since (F) is also more congruent with the stereotype stemming from the associated personality description than (F&T)). Researchers may be correct in assuming that this answer arose from logical considerations; however, there is no way to be sure of this. It is therefore important to identify whether these correct and logical answers are determined only by heuristic processing, or whether the participants are sensitive to the logic that underpins the conjunction rule.

The Comparison Conjunction Probability Judgement (CCPJ) Task

Following Villejoubert (2009; see also Curtis-Holmes & Evans, 2005), the experiments described in this thesis use a research methodology aiming to disentangle the influence of intuitive and representativeness considerations underpinning conjunction fallacies. This innovative experimental procedure allows the examination of the relative impact of heuristic and logical considerations on conjunction judgements. Instead of asking participants to make separate evaluations of the probability of a conjunction and the probability of a single event, this procedure asks them to judge whether statements comparing both types of probabilities are true or false (see Figure 2. 1). These statements vary in terms of their logicality (logical vs. illogical) and their compatibility with the representativeness heuristic (representative vs. unrepresentative). The four possible statement types are illustrated in Table 2. 1.

A. Classic Conjunction Fallacy Task	**B. Comparative Conjunction Probability Judgement Task**
Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations.	Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations.
Which one of the following statements is most likely? a. Linda is an active feminist. b. Linda is a bank teller and is an active feminist.	Is the following statement correct? Linda is more likely to be an active feminist than she is to be a bank teller and an active feminist. TRUE or FALSE

Figure 2. 1. Illustration of the classic conjunction fallacy task compared to the Comparison Conjunction Probability Judgement (CCPJ) task. *Note*. The left panel (A) shows the classic version of the Linda problem where participants rank order probabilities. The right panel (B) shows the CCPJ task where participants accept a statement containing both types of probabilities as true or reject it as false

Table 2. 1

Illustration of the four types of statements used to study the conjunction fallacy in CCPJ tasks.

	Conflict statements
R/I	Lynn is more likely to be a cashier and a feminist than she is to be a cashier.
U/L	Lynn is more likely to be a cashier than she is to be a cashier and a feminist.
	No conflict statements
R&L	Lynn is more likely to be a feminist than she is to be a cashier and a feminist.
U&I	Lynn is more likely to be a cashier and a feminist than she is to be a feminist.

Note. R/I: Representative but Logical; U/L: Unrepresentative but Logical; R&L: Representative and Logical; U&I: Unrepresentative and Illogical.

The CCPJ task makes it possible to manipulate whether or not a conflict is present in the probability statement. By presenting a series of no-conflict and conflict statements to participants with varying cover contents, it is possible to compute an index of their sensitivity to logic considerations. With no-conflict statements, both representativeness and logic points towards the same answer. For example, if a participant *accepts* a representative and logical (R&L) statement as true, they are answering in accordance with both representativeness and logic considerations. The same can be said for *rejecting* an unrepresentative and illogical (U&I) statement as false. Thus, the R&L statements should be easy to accept as true, while the U&I statements should be easy to reject as false. The conflict statements (i.e., containing conflict); however, should be more difficult to answer if people detect the conflict because representativeness and logic considerations point towards opposite answers. For example, if a participant accepts a representative but illogical (R/I) statement as true, they have transparently chosen the heuristic answer over the logical one. However, if they accept an unrepresentative but logical (U/L) statement as true then they chose the logical answer over the heuristic one. These conflict statements allow to test whether people are sensitive to the conflict between logic and representativeness. The CCPJ task allowed me to observe the effects of conflict on people's conjunction probability judgements, including their judgement latencies and eye movements.

The original CCPJ task (Villejoubert, 2009) was comprised of 12 scenarios. Two of these were the original Linda and Bill descriptions (Tversky & Kahneman, 1983). In addition, ten new descriptions were modelled after these original texts. The statements fell into the same four categories as the current CCPJ task: R&L, R/I, U/L, and U&I. Thus, there were three scenarios for each category of statement. The task was originally designed and administered in French, therefore it was translated into English for this program of research. Additional improvements were made by adding four extra scenarios to the task bringing the total number of scenarios to 16. This made it possible to design the study (and analyse it) across four blocks of four scenarios each. An example of a scenario is as follows: "Jesse is 28 years old. She enjoys meeting new people and traveling to unusual countries. She has excellent team building skills and is very athletic. In her spare time she likes to go hiking and play paintball." The scenario is then accompanied by a statement such as: "Jesse is more likely to be a soldier than a beauty salon owner and a soldier" (R&L statement; see Appendix 1 for the full CCPJ task).

Each trial presented a short person description followed by one of the four types of probability statements. Each description was designed so that the person presented appeared both as a highly representative member of a stereotypical category and highly unrepresentative of an atypical category, without openly mentioning that this individual belonged to either category. These descriptions were then tested in a pilot study where 20 students were asked to rate how shocked they felt when they saw two events listed together. For example, they were asked to rate on a scale from 0 (not at all shocked) – 9 (extremely shocked) how shocked they were to learn that Neil was both a “builder and a dance instructor”, or that Sarah was a “foreman and fortune teller”. The thumbnail descriptions that were rated highest on the “surprise factor” were included in the final CCPJ task. The final version of the task contained 16 trials. See Appendix 1 for the full CCPJ task used in this thesis. Each scenario was found in only one condition (i.e., the type of statement was not rotated through scenario). Participants saw each of the sixteen scenarios only once; however, the type of statement seen after each scenario differed across participants due to the design and randomisation of the CCPJ task, In other words, the Linda scenario was followed by the random selection of either an R&L, R/I, U/L or U&I statement. Plus, the order in which the scenarios were completed was randomised across participants.

Half of the 16 statements presented were logical (e.g., R&L and U/L) and the remaining half were illogical (e.g., the R/I and U&I, see Table 2. 1). Within each set of 8 statements, half were compatible with predictions based on representativeness (R&L and R/I statements) whereas the other half were incompatible with those predictions (U/L and U&I statements). Representativeness was determined by the short description presented before the statement and was manipulated by crafting each description so that the individual presented appeared both as a highly representative member of a given category and highly unrepresentative of an “odd” category, without openly mentioning that this individual belonged to either category. “R” statements always described the individual as more likely to belong to the stereotypical category whereas “U” statements always described the target as more likely to belong to the odd category. These “R” and “U” statements were then combined with the law of conjunction probability to be either logical (R&L, U&L) or illogical (R/I, U/I). In line with this law of probability, participants who judged the single clause as more likely to occur than the conjunction clause made a *logical* judgement, while those who judged the conjunction clause as more likely made an *illogical* judgement. The statement manipulation between representativeness and logicality afforded an insight into the cognitive underpinnings of people's probability judgements as the given answer pointed to the thinking style that was employed while making the judgement: either intuition or logic.

Each type of statement was presented four times in a random order. For all of the experiments, except the eye-tracking experiment, the CCPJ task was programmed using E-Prime and participants completed it from a desktop computer. For each CCPJ trial, all participants were first presented with a thumbnail description of a fictitious person and once they had finished reading it, they were instructed to press the spacebar in order to move onto the next part. This resulted in a probability statement appearing below the description. Participants read the statement and then responded "True" or "False" indicating whether or not they agreed with the statement. The programme recorded the time taken to read the description, the time taken to read the statement and make a judgement, the total test time from start to finish, as well as the responses given. The eye-tracking experiment was similar but had a few small differences. It was programmed in Matlab, and both the description and the statement were presented simultaneously for each trial. The software recorded all the variables mentioned above, as well as the number of refixations into the areas of interest (AoI), the refixation dwell times, the first run dwell times, and the total dwell times. The AoIs comprised of the description and the statement, and then the statement was divided up into several areas: "more likely", "single clause", "than" and "conjunction clause". For an example, see Figure 2. 2.

Lynn is 31 years old. She is single, outspoken, and very bright. She has a Degree in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Lynn is *more likely* to be *a feminist* *than* she is to be *a cashier and a feminist*. TRUE FALSE

Figure 2. 2. An example of a CCPJ trial as presented to participants in the eye-tracking study (Experiment 2). *Note*. AoIs are represented inside of asterisks; this is an example of an R&L statement.

No two CCPJ tasks in this program of research were identical. They all differed slightly according to the manipulations being applied (e.g., time pressure, ego-depletion, or cognitive load); thus, more specific details about the procedures used in each experiment will be provided in the following chapters.

Ethical Approval

For all five of the experiments conducted across this research project, ethical application was made to the Kingston University Faculty Research Ethics Committee. Experiment 1: “The Impact of Preferred Thinking Styles and Time Pressure on the Conjunction Fallacy”, was approved and granted permission to proceed on 25 April 2012. Experiment 2: “Eye-tracking Evidence for Logical Intuitions in Conjunction Probability Judgements”, was approved on 18 January 2013. Finally, experiment 3: “The Influence of

Ego Depletion and Impulsivity on the Conjunction Fallacy”, experiment 4: “The Influence of Spatial Storage Load and Cognitive Reflection on the Conjunction Fallacy”, and experiment 5: “The Effects of Time Pressure and Cognitive Reflection on Base Rate Neglect” were approved on 26 March 2014.

The procedures needed to conform to ethical scrutiny ensured no harm came to anyone who chose to participate in the studies. This included ensuring all participants were fully informed before consenting to participant in the study. Their participation remained anonymous and they could withdraw from the study at any point. Following the completion of the study, all participants were debriefed and reimbursed for their time and efforts. The participants were also offered the option of withdrawing their data from analysis at a later stage and were provided with a cut-off date for this. The studies were not of a stressful nature, nor were they likely to trigger a negative experience. All the signed consent forms and collected data were stored safely on Kingston University campus.

Chapter 3: Analysing Response Times and Eye Movements to Further Understand the Conjunction Fallacy

People typically use heuristic (i.e., intuitive) thinking to make many of their day-to-day judgments and decisions. For example, judging the quality of a product based on its price, or judging a job candidate based on his or her level of assertiveness. In fact, most of the "educated guesses" we make are founded on heuristics. It would be very impractical to write exhaustive lists or use decision trees every time we needed to make a judgment. Heuristic thinking is often assumed to involve fast decision strategies that save time and effort (Kahneman & Tversky, 1973, Kahneman & Frederick, 2005, Evans, 2008). It is considered to be an efficient and reliable process in most instances. Such instances would include when heuristic responses are congruent with logical principles. However, there are instances when heuristic responses clash with laws of logic and violate axioms of probability, and in these instances heuristic thought leads to erroneous, so-called "biased" judgments. In the situation of conjunction probability judgements, making these errors is termed committing *conjunction fallacies* (Tversky & Kahneman, 1983).

People often deviate from what are considered to be normative responses in judgement and decision-making tasks. A current debate in the literature asks the question: are people aware when their heuristic response conflicts with logical considerations? Default-interventionist models assume that heuristic processes are always activated first and that deliberate processes are activated only if necessary to intervene, correct or support heuristic reasoning (Evans, 2007; Glöckner & Witteman, 2010). Erroneous judgments result from the failure to detect the conflict between heuristic and logical considerations; which explains why heuristic responses are readily endorsed without scrutiny from the deliberative system (Kahneman & Frederick, 2002, Kahneman, 2011). Parallel-competitive models argue that people are able to detect the conflict but fail to inhibit the enticing heuristic response and replace it with a more deliberate logical response (Thompson, Turner, & Pennycock, 2011, Evans & Stanovich, 2013, Reyna & Brainerd, 2008, Sloman, 1996). Both of these models assume that logical answers emerge from slow, deliberative, and rule-based processes whereas heuristic answers arise from faster, intuitive, and associative processes. It is widely assumed that the "fast thinkers" who rely on heuristic thinking by default, are also prone to making biased judgments and committing fallacies (Tversky & Kahneman, 1983, Kahneman, 2011, Evans, 2003); however, few studies have empirically put this claim to the test. For example, in the original conjunction fallacy task (i.e., the Linda problem by Tversky and Kahneman, 1983), it is not possible to assess whether people's judgments are solely informed by heuristic considerations or whether they are also sensitive to logic considerations even if their final answer is congruent with a heuristic assessment.

The Conjunction Fallacy. The conjunction rule of probability states that the likelihood of two events occurring simultaneously (i.e., a conjunction) is always less probable than either one of the constituent events occurring alone. Conjunctions have an extensional nature, which means that the probabilities associated with the conjunction of two events occurring simultaneously (e.g., F&B) are included in the probability that event F, or event B, will occur individually (Tversky & Kahneman, 1983). The conjunction fallacy is observed when a participant assigns a higher likelihood to the conjunction of two events (i.e., F&B > B) than to one of the constituent events alone. The best-known demonstration of the conjunction fallacy presented a description of a woman named Linda to participants. Linda was described as being "31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations" (Tversky & Kahneman, 1983, p. 297). Eighty-five percent of participants believed it was more likely that Linda was a "bank teller and a feminist" than that she was simply a "bank teller". This violates the conjunction rule of probability because Linda's being a bank teller includes the possibility that she is also a feminist, so the conjunction cannot be more likely. See Figure 1. 2, pg.18 for the original rendition of the Linda and Bill tasks.

Tversky and Kahneman (1983) attributed the conjunction fallacy to the use of the representativeness heuristic when making conjunction judgements. Representativeness is an evaluation of the degree to which something corresponds with a stereotype or prototype (Tversky & Kahneman, 1983). The representativeness heuristic is a cognitive tool that uses past knowledge to estimate probabilities (Kahneman & Tversky, 1973). Although this heuristic is fast, and at certain times efficient at making quick, snap or gut instinct judgements, when used to make conjunction probability judgements it leads to errors and biased judgments. Participants seemingly assess how similar Linda is to the stereotypes of bank tellers and feminists.

Does intuition always mislead people's judgements? While there is a consensus supporting the assumption that erroneous judgements originate from rapid considerations and logical judgements arise from slower, deliberative considerations; there is very little empirical evidence to support this theoretical claim. An accepted assumption from most dual process accounts of the conjunction fallacy is that heuristic considerations (i.e., derived from representative considerations) arise from intuitive processes, while logical considerations (i.e., derived from statistical rules of probability) arise from deliberative processes. Controversially, the recent emergence of the logical intuitions account states that logical considerations might in fact stem from intuitive processes (De Neys, 2012), while others believe that slower deliberative reasoning can be aligned with reflective heuristics (Villejoubert, 2009; Howarth, Handley & Walsh, 2016). Thus, without empirical evidence it is difficult to define which system of reasoning drives intuition and which drives logic.

Recent research by Evans, Dillon and Rand (2015) has found that extreme social decisions were fast, but not intuitive. The authors caution other researchers in experimental psychology and economics not to interpret slow reaction times as evidence of reflective thinking. In support of this, a recent alternative theoretical position calls for distinguishing heuristic-biased thinking from intuitive processes. Heuristic thinking may lead to biased judgements, but it is unclear of whether this is happens while deliberating or reflecting. Contrary to the traditional two-systems approach, there is empirical evidence suggesting that intuitive processes are not necessarily biased. Indeed, people can be fast and logical and they can intuitively and implicitly detect the conflict between logical and heuristic considerations (Evans, Dillon & Rand, 2015; De Neys, 2012; Villejoubert, 2009).

The idea that people can have logical intuitions is supported by recent work on conflict detection in reasoning, which explores whether or not people detect that they are biased when making a judgement. Traditionally the default interventionist account has accepted that fast, automatic and intuitive processes stem from System 1 processing, which is susceptible to heuristic biases and result from the failure to detect the conflict between deliberative and intuitive considerations. While conflict detection is assumed to only be available to deeper, slower, more effortful deliberative cognitive processing (e.g., see Kahneman, 2011). Parallel competitive accounts; however, state that people are very good at detecting conflict between the two systems of reasoning, but they make errors due to the failure to inhibit the convincing heuristic answer (Sloman, 1996).

The following studies used a variety of measures, including response latencies (e.g., Bonner & Newell, 2010; Villejoubert, 2009), self-report confidence ratings showing how confident they were in their judgements (De Neys, Cromheeke & Osman, 2011) and eye tracking (Ball, Phillips, Wade & Quayle, 2006), to examine if people were sensitive to violations of statistical laws of logic or, in other words, whether they were able to detect the conflict between the heuristic and logical considerations. Typically, conjunction fallacy tasks presented to the participants contained both congruent (i.e., no conflict) and incongruent (i.e., containing conflict) statements. Take the Linda problem for example; a congruent statement would present the most probable option as the one that is both representative of the stereotype, and also in line with the law of probability. For example, participants can be asked to "rank which is more likely to occur: Pr(F) or Pr(F&B)". Ranking is relatively straightforward due to there being no conflict between the logical (one event is more likely than a conjunction of events, so Pr(F) > Pr(F&B)) and the heuristic answer (Linda *looks like* a feminist, so Pr(F) > Pr(F&B)). However, things get more complicated when the task is to "rank which is more likely to occur: Pr(B) or Pr(F&B)". This proves more difficult to solve because there is conflict between the logical and heuristic considerations. From a logical point of view, one event is more probable than a conjunction so Pr(B) should be ranked as more probable than Pr(F&B) but Linda's portrait does not resemble that of a bank teller, so Pr(F&B) may seem more probable than Pr(B). The results from the conflict detection studies showed that people typically needed more time to solve the conflict compared to the control versions of the reasoning problems. They also rated their confidence levels as lower than the control versions, suggesting that they were sensitive to the fact that their answers were inconsistent with the logical rules (De Neys, Cromheeke & Osman, 2011). It is interesting to note, that in some instances this conflict detection caused further deliberation resulting in more logical answers; however, this was not consistent across all studies and in some instances conflict detection did not transpire in people making correct judgements. The research suggests that all reasoners (both biased and unbiased) are able to detect conflict between intuitive and deliberate assessments; thus, failure to provide an unbiased answer is not due to a lack of conflict detection (De Neys, 2015). Previous research has shown that people perform better on incongruent trials (i.e., when conflict is present) than congruent ones (i.e., when conflict is absent) because the detection of two possible answers activates a deeper processing of the problem (Tremoliere & De Neys, 2014). The effect that conflict has on the logicality of probability judgements will be addressed further across this program of research.

Experiment 1: The Impact of Preferred Thinking Styles and Time Pressure on the Conjunction Fallacy

The logical-intuitionist account suggests that people's intuition does not always mislead them. It can sometimes reflect sensitivity to the logic of sets that underpins the conjunction rule. This account assumes that initial, fast judgements are logical but are subsequently overridden by slower, deliberative heuristic thinking. In other words, it is the logical answer that comes first, and not the heuristic answer as previously believed. There is evidence to show that people can make fast and logical judgements, and they can intuitively detect the conflict between logical and heuristic considerations (Saxe, 1988; De Neys, 2014; Villejoubert, 2009). Villejoubert (2009) provided initial evidence for this idea; however, the study was never replicated. There were also a few limitations to the study, such as the data was collected online, thus time pressure was not controlled. The current study was intended to replicate the findings of Villejoubert (2009), shed light on the debate surrounding logical intuitions and address the following point of contention: whether people can have logical intuitions in conjunction probability problems or whether they are simply applying the representativeness heuristic blindly.

The Present Experiment. An experiment was designed to test the hypothesis that individuals who engage more naturally in effortful, deliberate thinking are also more prone to produce illogical probability judgements if they have time to deliberate, because it is the slower, more deliberative form of thinking that may be more vulnerable to prior beliefs and non-logical heuristics. A methodology adapted from Villejoubert (2009) was employed to investigate whether people are "cognitive misers" who fail to reason in line with the tenets of logic because they either lack the cognitive ability or the motivation to do so. This study combined the new adapted Comparison Conjunction Probability Judgement (CCPJ) task with the Rational Experiential Inventory (REI-40; Pacini & Epstein, 1999), a 40-item scale designed to measure people's propensity to think rationally or experientially.

Rational-Experiential thinking styles and Dual-Processes. The REI is designed to measure two independent processing modes by combining two scales: the Need for Cognition Scale (NFC; Cacioppo & Petty, 1982) with the Faith in Intuition Scale (FI; Epstein, et al., 1996). The NFC is the tendency for an individual to engage in and enjoy thinking (Cacioppo & Petty, 1982), while the FI scale reflects how much trust a person has in his or her own intuition (Alos-Fereer & Hugelschafer, 2012). In other words, it is an interplay between the two thinking styles: System 1 and System 2. It has been found that high scores on the FI scale is positively correlated with committing the conjunction fallacy (Toyosawa & Karasawa, 2004). Alos-Fereer and Hugelschafer (2012) also found that intuitive decision makers relied more heavily on the representativeness heuristic; however, they cautioned that FI "should not be equated with heuristic processing, since intuitive people seem to be susceptible to particular heuristic cues and not to others" (p. 187). By contrast; however, the logical intuitionist account would predict that faith in intuition may instead result in more logical answers in the particular context of the conjunction fallacy where the probability relationship between single and conjunctive events may be implicitly learnt from experience. With this in mind, and in line with the assumption that intuitive thinking can display the tenets of logic (i.e., the logical answer comes first), this experiment aimed to examine whether individuals who relied more heavily on experiential thinking would also show more logicality when making conjunction judgements, especially when conflict was present in the statement.

Time Pressure and Judgement and Decision Making. Time pressure was employed as the task manipulation as a means to obstruct deliberative thinking in half of the participants. The other half were in the control condition and were not subjected to any time limitations. It has been traditionally accepted that time pressure impairs participants' ability to reason about the details of the situation and thus promotes intuitive thinking (Capraro & Cococcioni, 2016; Sloman, 1996; Kahneman, 2011; Evans & Stanovich, 2013; Stanovich & West, 1998). Classically, time pressure is believed to increase the likelihood that people use readily available strategies, such as heuristics and System 1 thinking. "The logic being that a key feature of intuitive processes is that they can be executed more quickly than deliberative processes" (Krajbich et al., 2015, p. 2). Without the luxury of having time available, individuals will rely more heavily on preexisting strategies (i.e., heuristics). But what if these preexisting strategies have been learnt implicitly by simply existing in the world around us? Could it be possible that we have acquired knowledge of statistical laws by simply interacting with our surroundings? Take Saxe's (1988) research as an example, he found that Brazilian street children who were uneducated and never attended school, were able to perform complex mathematical equations when selling their candies on the streets. Indeed, they outperformed children of the same age who attended school and were taught mathematics. These children had implicitly acquired knowledge of algebra through a necessity to survive. In accordance with the assumption that the heuristic answer comes first and can be logical, while the deliberative answer is susceptible to biases, this experiment aimed to test whether participants who were put under extreme time pressure would be less likely to favour the heuristic answer; while given time to deliberate they would succumb to the representativeness heuristic to inform their answers.

Method

Participants. A power analysis indicated that a total sample size of 70 students would be sufficient to detect effects of a medium size effect (d = .5), as reported in Villejoubert (2009), with 1 – b = .80 and a = .05. The final sample comprised of 101 students ($n_{unlimited}$ = 50; $n_{limited}$ = 51), ranging in age from 18-56 years (M = 24.16, SD = 8.07). Eighty were females

(79.2%) and 21 males (20.8%). Most of the students were undergraduates (*n* = 68, 67.3%) and the remaining 32 students were postgraduates (31.7%). Almost all of participants in the sample were enrolled on Psychology modules (*n* = 86, 85.1%), however 14 students (13.9%) were enrolled on other courses. Fifty-eight of the students were native English speakers (57.4%), while the remainder of the sample spoke English as a second language (*n* = 42, 41.6%).

Apparatus and materials. Participants were required to complete a series of 16 Comparative Conjunction Probability Judgement (CCPJ) tasks. Each task presented a short person description followed by one of four types of probability statements (see Figure 3. 1 for a screenshot of a trial presented to the participants in the unlimited time condition). Half of the 16 statements presented were logical (e.g., the Representative & Logical and Unrepresentative but Logical statements) and the remaining half were illogical (e.g., the Representative but Illogical and Unrepresentative & Illogical statements). Within each set of 8 statements, half were compatible with predictions based on representativeness (R&L and R/I statements) whereas the other half were incompatible with those predictions (U/L and U&I statements; see Table 3. 1 for an illustration of the four types of statements used to study the conjunction fallacy). Representativeness was determined by the short description presented before the statement and was manipulated by crafting each description so that the individual presented appeared both as a highly representative member of a given category and highly unrepresentative of an "odd" category, without openly mentioning that this individual belonged to either category. R statements always described the individual as more likely to belong to the stereotypical category whereas U statements always described the target as more likely to belong to the odd category. R and U statements were also defined so that they could nevertheless be logical (R&L, U&L) or illogical (R/I, U/I). In line with the conjunction law of probability, participants who judged the single clause as more likely to occur than the conjunction clause made a *logical* judgement, while those who judged the conjunction clause as more likely made an *illogical* judgement. The statement manipulation between representativeness and logicality afforded an insight into the cognitive underpinnings of people's probability judgements as the given answer pointed to the thinking style that was employed while making the judgement: either intuition or logic. See appendix A for the 16 descriptions and their accompanying statements.

Table 3. 1

Illustration of the four types of statements used to study the conjunction fallacy in CCPJ tasks

Key	Statement
	Conflict statements
R/I	Lynn is more likely to be a cashier and a feminist than she is to be a cashier.
U/L	Lynn is more likely to be a cashier than she is to be a cashier and a feminist.
	No conflict statements
R&L	Lynn is more likely to be a feminist than she is to be a cashier and a feminist.
U&I	Lynn is more likely to be a cashier and a feminist than she is to be a feminist.

Note. R/I: Representative but Logical; U/L: Unrepresentative but Logical; R&L: Representative and Logical; U&I: Unrepresentative and Illogical.

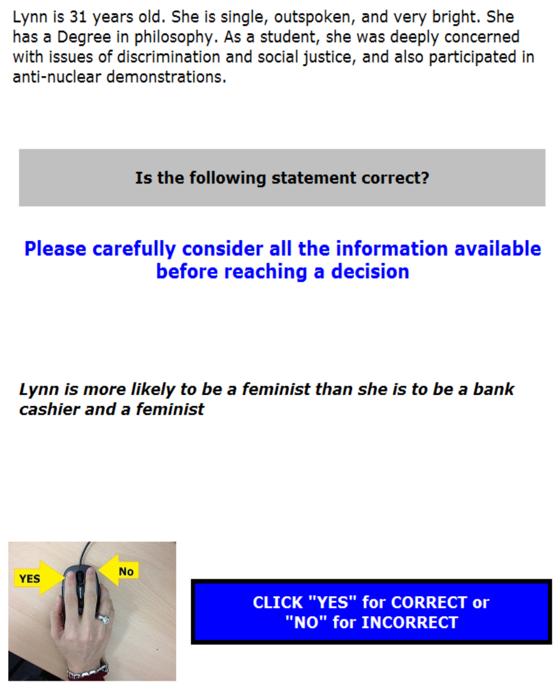


Figure 3. 1. Screen shot of the Lynn scenario in the CCPJ task as seen by the participants. The statement depicted is representative and logical (RL). This trial was in the unlimited time condition

Each type of statement was presented four times in a random order using a total of sixteen different thumbnail descriptions and events. The experiment was programmed using E-Prime and participants completed it from a desktop computer. For each CCPJ task, all participants were first presented with a thumbnail description and instructed to click the left button on the mouse once they had finished reading it. This resulted in a probability statement appearing below the description. Participants then left clicked the mouse for “True” or right clicked the mouse for “False” indicating whether or not they agreed with the statement. The programme recorded the time taken to read the description, the time taken to make a judgement, the total test time from start to finish, and the answers given. Following this, participants completed an electronic version of the Rational-Experiential Inventory (REI-40) questionnaire which was composed and executed in Qualtrics.

The purpose of the REI-40 (Pacini & Epstein, 1999) was to assess preferences for information processing between the two thinking styles. The REI scale distinguishes between 2 cognitive styles: A rational style, measured by an adapted NFC scale (Cacioppo & Petty, 1982), emphasises a conscious, analytical approach; and an experiential style, measured by the FI scale (Epstein, et al., 1996), emphasises a pre-conscious, affective, holistic approach. The scale includes 40 items and uses a 5-point ratings (1 = definitely not true of myself to 5 = definitely true of myself). There are 4 sub-scales with 10 items each: Rational Ability (RA) is the ability to think logically and analytically; Rational Engagement (RE) is one's reliance on and enjoyment of thinking in an analytical, logical manner; Experiential Ability (EA) is the ability with respect to one's intuitive impressions and feelings; and Experiential Engagement (EE) is one's reliance on and enjoyment of feelings and intuitions in making decisions. Sample items include: "I have a logical mind" (RA), "I enjoy intellectual challenges" (RE), "I believe in trusting my hunches" (EA), and "I like to rely on my intuitive impressions" (EE). For the complete REI-40 scale see Appendix 2.

Design and Procedure. This study used a 2 x 2 x 4 mixed design with the time condition (unlimited time vs. time pressure) as a between-subject factor, and statement conflict (absent vs. present) and block (four trial blocks) as within-subject factors. The 16 statements were comprised of 4 blocks, and each block had 4 statements (e.g., block 1 = RL1, RI1, UL1, UI1; block 2 = RL2, RI2, UL2, UI2, etc.) Two random orders of presentation of the trials were used to minimise confounding variables. The order of the trials in version 1 was reversed for version 2 (e.g., an R&L trial in version 1 became a U&I trial in version 2).

In addition to this, the trials for both versions were programmed to be presented in a random order for each participant.

Each vignette was presented on its own and participants had as much time as they needed to read it. They then pressed the spacebar to continue, which caused the statement to be presented. Participants were randomly allocated to one of two experimental conditions: the unlimited time or the time pressure condition. Those in the unlimited time condition were given as much time as they wished to answer the statements. They received instructions that stated: “Take all the time you need to read each description and consider the information carefully before reaching a decision” and were prompted throughout the experiment affirming that they may take as long as they wanted to reach their decision. Prompts included “please carefully consider all the information available before reaching a decision” and “please take all the time you need to read the following description”. Participants in the time pressure condition received the following instructions: “Read the description carefully BUT hurry to record the first snap answer that comes to your mind”. Additionally, they were incentivised to think quickly by beginning the test with £4 in “the bank” and every time they failed to answer within 8 seconds they lost 25p from the “bank” (25p x 16 CCPJ tasks = £4). They received feedback after each CCPJ task informing them how long they took in seconds to respond and what their remaining balance was. Throughout the experiment participants received prompts that stressed the urgency to make fast judgements and decisions, for example: “remember to respond as fast as you can!” and “please HURRY and record the first answer that comes to your mind!” (see Figure 3. 2 for a screenshot of a trial presented to participants in the time pressure condition). To ensure no data were missing there was no actual time limit, only the impression of one. No participants exceeded the time limit of 8 seconds ($M_{limited}$ = 4.4 seconds, SD = 1.2, Min = 1.4, Max = 7.8). Regardless of their final money balance, all participants were paid £4 for their co-operation, but this was unknown to them until after they had completed the task. The experiment took place in a psychology laboratory on the Kingston University campus. Participants gave their informed consent before anonymously completing the CCPJ task. Participants also completed an electronic version of the REI-40 questionnaire and gave self-report feedback as to how they solved the CCPJ tasks. Following this, they were debriefed and paid.

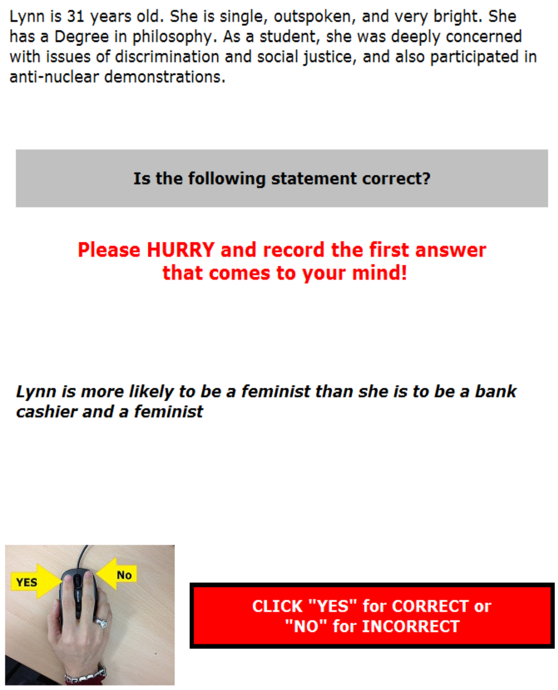


Figure 3. 2. Screen shot of the Lynn scenario in the time pressure condition. The statement depicted is representative and logical (RL)

Results

Demographics and outliers. A total of 101 Kingston University students participated in the experiment (N = 101, *n_limited* = 51, *n_unlimited* = 50). To examine the extent to which individual responses may be consistently aligned with the logical rule or the representativeness heuristic, the sample was screened for participants who always accepted R&L and U/L and rejected R/I and U&I statements (strictly logical), as well as participants, if any, who always *rejected* R&L and R/I statements and *accepted* U/L and U&L statements (strictly odd), to gauge the extent to which the overall results could be influenced by consistent responding. No reasoners exhibited choices which were always logical, nor always at odds with heuristic considerations, thus the sample size remained at N = 101.

Robustness Check. Following the procedures outlined by Meade and Craig (2012), an overall "flag score" was computed to identify careless responders who were identified as outliers on a number of criteria. Three participants in total were removed from the sample. It was found that removing these three participants who were flagged more than once, made no difference to the results of the analyses. Thus, the results are robust even when the outliers are retained in the sample. For this reason, the detailed results for the entire sample (N = 101) will be reported. See Appendix C for a full version of the robustness check results.

Manipulation Checks. On completion of the study each participant was asked to report how much time pressure they had experienced during the task. This was recorded on a 10-point scale from 0 – 9 (0 = "no time pressure", 9 = "extreme time pressure"). Manipulation checks revealed that participants in the limited time condition reported feeling under more pressure than those in the unlimited time condition ($M_{unlimited}$ = 1.72, SD = 2.33 vs. $M_{limited}$ = 5.90, SD = 2.18; $t(99)$ = -9.33, p < .001, Cohen's d = 1.86).

The total time spent completing the study was compared across the two time conditions. It was found that the participants in the limited time condition completed the study significantly faster than those who had no time pressure ($M_{unlimited}$ = 471.75 seconds, *SD* = 162.84 vs. $M_{limited}$ = 341.29 seconds, *SD* = 87.93; $t(99) = 5.02$, $p < .001$, Cohen's $d = 1.00$).

Following this the average reading time for the scenarios was calculated for each participant and compared across the two time conditions. There was no difference in reading times between the unlimited and limited time conditions ($M_{unlimited}$ = 11.84s, *SD* = 3.73 vs. $M_{limited}$ = 12.36s, *SD* = 4.23; $t(99) = -.65$, $p = .517$, Cohen's $d = .13$).

The final manipulation check examined whether there were any differences in judgement latencies across the two time conditions. In other words, the average time each participant took to read the statements and make their judgements. Results confirmed that participants who were put under time pressure did make their judgements significantly faster than those who had unlimited time to complete the study ($M_{unlimited}$ = 15.54s, *SD* = 7.64 vs. $M_{limited}$ = 4.38s, *SD* = 1.21; $t(99) = 10.30$, $p < .001$, Cohen's $d = 2.05$).

Manipulation checks confirmed that participants in the limited time condition were under time pressure to respond quickly but were not pressured to read the task description faster than participants in the unlimited time condition. The fact that every participant spent the same amount of time reading the information before making their judgements means that any differences in reasoning cannot be attributed to differences in reading latencies. Participants in the limited time condition were not only faster in making their judgements, their overall time to complete the task was faster too and they reported feeling under more time pressure than those in the unlimited time pressure condition. This suggests that the time pressure manipulation was successful.

Conflict Sensitivity. To examine conflict sensitivity, a heuristic score was computed for each trial to assess whether the participants were sensitive to the conflict between heuristic considerations (i.e., representativeness) and logical considerations when making their judgements in conflict trials. Responses were coded as heuristic when participants accepted representative statements (R&L, R/I, see Table 2. 1, p. 74) or rejected unrepresentative statements (U&I, U/L). From these heuristic scores, I was able to calculate a total score for the conflict statements and for the congruent statements across each of the four blocks. The scores were maintained across the blocks to explore whether any order effects or learning effects were present in the data. Heuristic scores theoretically ranged from 0% (never accepted R&L or R/I statements nor ever rejected U/L or U&I statements) to 100% (always accepted R&L or R/I statements and always rejected U/L or U&I statements) and were analysed using a 2 (time condition) x 2 (conflict condition) x 4 (trial blocks) mixed analysis of variance (ANOVA) with repeated measures on the last two factors.

The results revealed that the heuristic scores of the participants did not change across trial blocks ($F(3, 297) = .82$, $p = .486$, $\eta_p^2 = .008$). In addition, neither the interaction between trial blocks and time pressure condition ($F(3, 297) = 1.19$, $p = .315$, $\eta_p^2 = .012$) nor the interaction between block trials and conflict ($F(3, 297) = 1.60$, $p = .191$, $\eta_p^2 = .016$) were significant. There was also no significant three-way interaction between block, conflict and time pressure conditions, $F(3, 297) = .362$, $p = .780$, $\eta_p^2 = .004$. This suggests that the order of the trials presented in each block had no effect on the logicality of the judgements produced. In other words, there were no practice or learned effects. Participants did not become more sensitive to the conflict between logical and heuristic considerations as they completed more trials.

Participants were more likely to respond heuristically in the absence of conflict between heuristic and logical considerations, $M_{no\_conflict} = 86.67$, $SD = 15.03$, 95% CI [83.77, 89.56] vs. $M_{conflict} = 78.67$, $SD = 21.38$, 95% CI [74.77, 82.58], $F(1,99) = 13.62$, $p < .001$, $\eta_p^2 = .12$. Heuristic thinking was also influenced by time; participants who were under time pressure to make a judgement were overall less likely to rely on heuristic considerations, $M_{unlimited} = 88.63$, $SD = 11.41$, 95% CI [84.81, 92.44] vs. $M_{limited} = 76.72$, $SD = 15.42$, 95% CI [72.94, 80.49], $F(1,99) = 19.41$, $p < .001$, $\eta_p^2 = .16$. There was a significant interaction between conflict and time pressure, $F(1, 99) = 4.80$, $p = .031$, $\eta_p^2 = .05$. As shown in Figure 3. 3, participants were less heuristic in their judgements when conflict was present in the statements, but this effect was even more noticeable when they were put under time pressure.

A simple main effects analysis showed that there was no difference in heuristic scores between the conflict and no conflict statements when time was unlimited *(p* = .294*)*; however, when put under limited time pressure this difference became significant ($p < .001$).

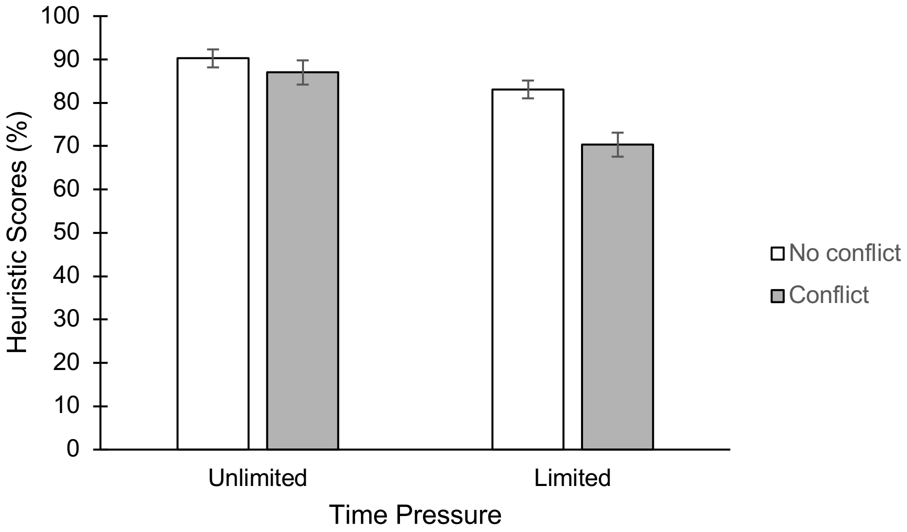


Figure 3. 3. Rate of heuristic responding as a function of the conflict between heuristic and logical considerations and as a function of response time condition. *Note*. Error bars represent standard errors. The per-cell sample size was $n = 50$ for unlimited time, and $n = 51$ for limited time

Judgement Latencies. Response latencies were analysed to assess which responses showed sensitivity to conflict between heuristic and logical considerations. Assuming that people who always respond with a heuristic judgement under conflict use a different strategy overall, I first screened and excluded them, so I could compare latencies for heuristic and logical answers under conflict ($n = 35$, 34.7%). In line with previous research (Pennycook, Fugelsang, & Koehler, 2012; De Neys & Glumicic, 2008) three average latency scores were computed for each of the remaining participants: the average response latency for congruent statements (R&L and U&I statements), and for the conflict statements (R/I and U/L statements), we computed the average response latency for heuristic responses, as well as the average response latency for logical responses. To compute the three latency scores, we first selected latencies for logico-heuristic answers in non-conflict trials (i.e., latencies to either accept or reject R&L and U&I statements). Next we selected the latencies for participants who responded heuristically to conflict trials (i.e., latencies to either accept R/I statements or reject U/L statements). Finally, we selected the latencies for participants who responded logically to conflict trials (i.e., latencies to either reject R/I statements or accept U/L statements). Latencies were converted to $\log^{10}$ prior to analysis to normalize the distribution and subjected to a 3(judgment type: no-conflict vs. conflict-heuristic vs. conflict-logical) x 2(time condition: unlimited vs. limited) mixed ANOVA with repeated measures on the first factor.

As expected, there was a significant main effect for time pressure condition with latencies being faster in the limited time condition than the unlimited condition, $M_{unlimited}$ = 19.00s, SD = 10.62, 95% CI [16.36, 21.63] vs. $M_{limited}$ = 4.38s, SD = 1.37, 95% CI [2.39, 6.37], $F(1, 64) = 171.04$, $p < .001$, $\eta_p^2 = .728$. There was also a main effect of judgement type with participants taking longer to make logical choices on conflict statements (i.e., rejecting R/I statements or accepting U/L statements; $M_{no\_conflict}$ = 9.97*s, SD* = 4.43, 95% CI [8.79, 11.15] vs. $M_{conflict\_heuristic}$ = 10.90s, SD = 6.04, 95% CI [9.25, 12.54] vs. $M_{conflict\_logical}$ = 14.20s, SD = 9.47, 95% CI [11.56, 16.85], $F(2, 128) = 8.99$, $p < .001$, $\eta_p^2 = .123$). There was a significant interaction between time pressure condition and judgement type, $F(2, 128) = 6.04$, $p = .003$, $\eta_p^2 = .086$. As Figure 3. 4 illustrates, when given unlimited time to respond, participants took longer to respond logically to conflict statements compared to heuristic responses to conflict statements, or to no-conflict statements. Under severe time pressure, however, these differences in judgement latencies disappeared.

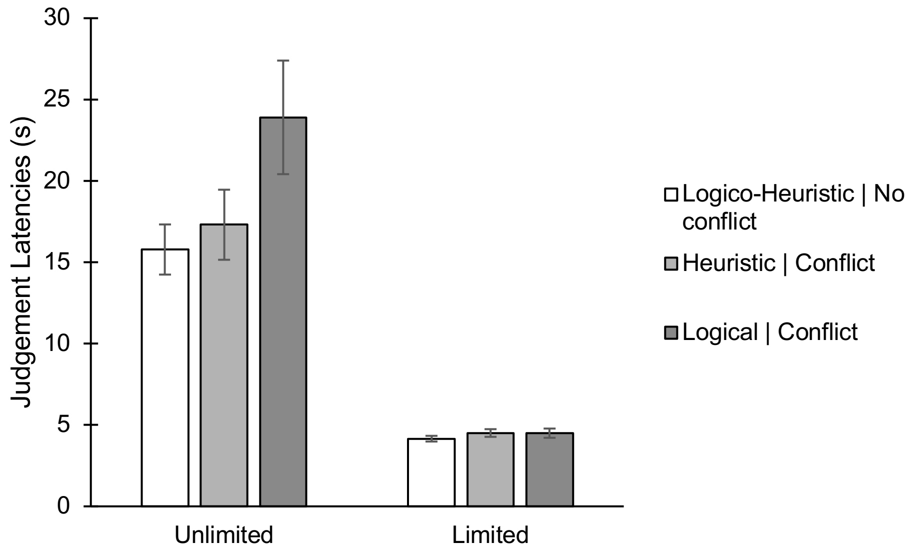


Figure 3. 4. Judgement latencies as a function of judgement type and time condition. *Note*. Error bars represent standard errors. The per-cell sample size was $n = 24$ for unlimited time, and $n = 42$ for limited time

A simple main effects analysis showed that when given unlimited time to make a judgement, participants took significantly longer to respond logically to conflict statements than to no conflict statements ($p < .001$), they also took significantly longer to respond logically to conflict statements than heuristically to conflict statements ($p = .007$). Interestingly, there was no difference in response times between no conflict statements and heuristic responses to conflict statements. ($p > .439$). Finally, when put under extreme time pressure, all these significant differences in judgement latencies disappeared. There was no difference in judgement times between the three response type conditions in the limited time condition (all *p's* > .05).

Preferred Thinking Modes and Logical Sensitivity. To further understand the origin of the detection of conflict in R/I and U/L statements, individual differences in thinking mode were evaluated to see whether they would predict whether individuals would favour logical choices (i.e., rejecting R/I statements and accepting U/L statements). If, as suspected, sensitivity to logic is intuitive, people who score higher on the experiential measure should be logical in their judgements on the CCPJ task. Responses to the REI scale were analysed using the Two-Step Cluster procedure from IBM SPSS 23.0, with 4 inputs: Rational Ability (average score from items 1, 4, 8, 13, 14, 17, 25, 27, 30, 39), Rational Engagement (average score from items 2, 6, 10, 16, 20, 26, 28, 32, 33, 40), Experiential Ability (average score from items 3, 5, 18, 19, 21, 34, 35, 36, 37, 38), and Experiential Engagement (average score from items 7, 9, 11, 12, 5, 22, 23, 24, 29, 31). To minimise order effects, cases were randomly ordered and cluster solutions were replicated using cases sorted in different random order to confirm their stability.

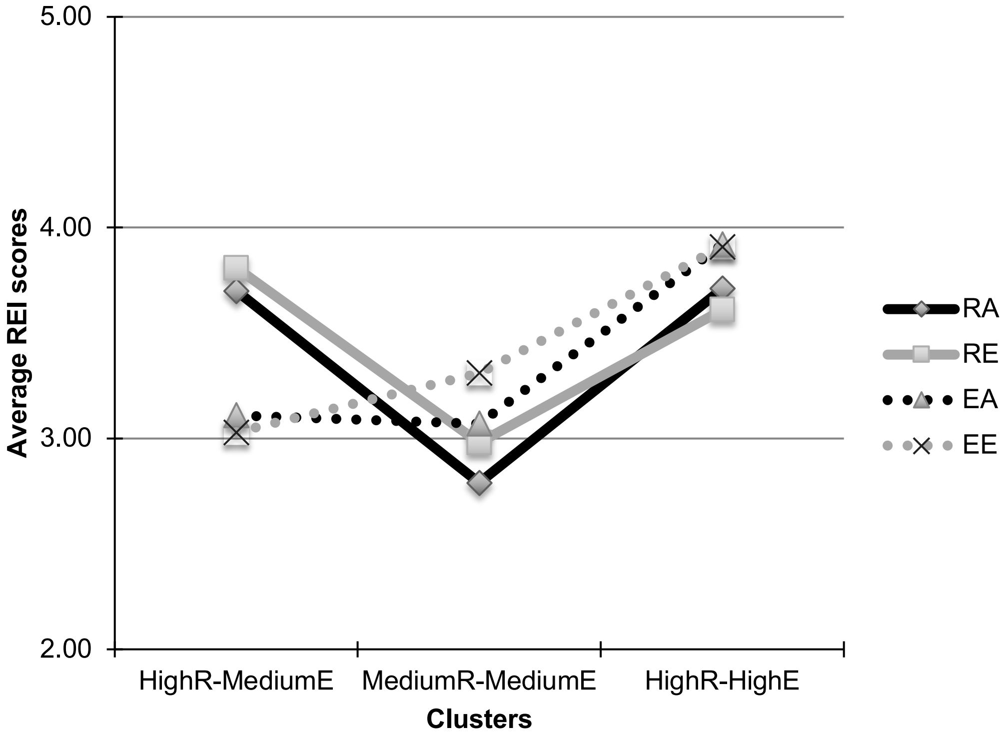


Figure 3. 5. Mean REI-score profiles for the three types of thinking mode dispositions.

The solution was stable and identified three clusters with an average silhouette measure of cohesion and separation of 0.40, suggesting a “fair” clustering solution. The ratio of sizes (largest cluster to smallest cluster) is 1.87. This number should be lower than 2 so this result is good. The first cluster profile is the largest (*N* = 43, 42.6%). As outlined in Figure 3. 5, it is characterised by participants who scored high on the RA and RE scales, but average on the EA and EE scales. Participants in this group view themselves as being far more rational than experiential. This cluster was labelled “highR-mediumE”. The second cluster profile was the smallest (*N* = 23, 22.8%). It is characterised by participants who scored average on the RA, RE, EA and EE scales. These participants did not rate themselves as being particularly rational or experiential. This cluster was labelled “mediumR-mediumE”. The third and final cluster profile (*N* = 35, 34.7%) is characterised by participants who scored high on the EA and EE scales, and high to average on the RA and RE scales. These participants rate themselves as both highly rational and experiential. This cluster was labelled "highR-highE". A logical score was computed for each person that showed the percentage of time they responded logically to the eight conflict trials. This logical score was subjected to a 2(time pressure: unlimited time vs. time pressure) x 3(REI profile: highR-mediumE vs. mediumR-mediumE vs. highR-highE) between subjects factorial analysis of variance (ANOVA). There was a significant main effect of time pressure with participants making more logical judgements when put under severe time pressure, $M_{unlimited\_time}$ = 11.09%, *SD* = 13.83, 95% CI [5.86, 16.33], $M_{time\_pressure}$ = 29.66%, *SD* = 22.23, 95% CI [24.34, 34.99], $F(1,95) = 24.37$, $p < .001$, $\eta_p^2 = .20$. There was also a borderline significant main effect of cluster profile; the mediumR-mediumE cluster performed the most logically in their judgements, followed by the highR-mediumE, while the highR-highE cluster performed the least logically of the three groups ($M_{mediumR\text{-}mediumE}$ = 26.66%, *SD* = 23.81, 95% CI [19.11, 34.21], $M_{highR\text{-}mediumE}$ = 19.98%, *SD* = 20.72, 95% CI [14.43, 25.54], $M_{highR\text{-}highE}$ = 14.49%, *SD* = 16.54, 95% CI [8.36, 20.63], $F(2,95) = 3.10$, $p = .05$, $\eta_p^2 = .06$. There was no significant interaction between time pressure and cluster profile, $F(2, 95) = 0.93$, $p = .400$, $\eta_p^2 = .012$. In other words, being put under severe time pressure, or no time pressure at all, made no difference to how people perceived their own thinking styles to be.

Post hoc comparisons using the Bonferroni correction indicated that the mean score for the mediumR-mediumE cluster (*M* = 26.66%, *SD* = 23.81) was significantly different than the highR-highE cluster (*M* = 14.49%, *SD* = 16.54). However, the highR-mediumE cluster (*M* = 19.98%, *SD* = 20.72) did not significantly differ from the mediumR-mediumE and highR-highE clusters.

Taken together, these results suggest that preferred thinking modes and time pressure do have an effect on people's logicality when making conjunction probability judgements. Specifically, these results suggest that when people are put under severe time pressure they make more logical judgements than when given unlimited time to make their judgements. Additionally, people who rated themselves as medium on both the rational and experiential scales were more logical in their judgements than those people who rated themselves as high on both the scales. However, it should be noted that there was no interaction between preferred thinking style and time pressure. Thus, the effect of preferred thinking style on logicality was not exaggerated under time pressure.

Discussion

Support of Hypotheses and Interpretation of Results. The experiment reported here examined the impact of time pressure on the sensitivity to the conflict between heuristic and logical assessment in social judgements under uncertainty. The results support the hypotheses that 1) people are readily able to detect the conflict between heuristic and logical considerations, however there was not enough evidence to convincingly support the hypothesis that 2) people can have logical intuitions in conjunction probability problems, they are not simply applying the representativeness heuristic blindly. Finally, there was no support for the hypothesis that 3) people are able to have heuristic reflections, in other words that the heuristic bias enters reasoning during deliberation.

Similarly to the research of De Neys and Glumicic (2008), the participants showed sensitivity to the conflict between representativeness and logic (i.e., they were sensitive to the differing answers produced by intuition and deliberation). Statements that contained conflict were associated with lower rates of heuristic judgements as well as longer judgement latencies, despite time pressure condition. This suggests that conflict detection happens readily and supports the ideas of Sloman (1996) and Epstein (1994) who claim that people experience a struggle between the two systems of thought and that human conflict monitoring is in fact very effective. Kahneman (2002) and Kahneman and Frederick (2005) state that conflict monitoring is quite lax and people make most decisions by relying on heuristic beliefs without taking logical considerations into account. In other words, errors occur because of a failure to detect conflict. There is support for this default interventionist account in Experiment 1; however, only in the judgement latencies of the participants in the unlimited time condition. Participants answered statements containing no-conflict in the same amount of time as they responded heuristically to statements containing conflict. This finding suggests that the representative (i.e., heuristic) answer comes quickly and easily to people, in fact, it suggests a similarity between the manner in which people respond to no-conflict statements and when they respond heuristically to conflict statements. Thus, it would appear from their judgement latencies, that in the unlimited time condition people were not always able to detect the conflict between heuristic and deliberate considerations. Instead they were happy to trust their immediate, intuitive judgement. Although, the heuristic response rate was high, there were some participants who answered the conflict statements logically, but they also took significantly longer to respond logically compared to their faster no-conflict and conflict-heuristic responses. The way the participants responded to the CCPJ task in the unlimited time condition was textbook default interventionist account: there was no evidence of conflict detection in their judgement latencies, plus for a few participants who were willing or able to apply their cognitive resources they were able to reach a logical answer in line with the conjunction rule, but this took significantly longer to achieve. Suggesting that System 1 is fast and heuristic, while System 2 is slower and more logical. However, when put under severe time pressure this effect completely disappeared and there were no differences found in the judgement latencies between any of the three judgement types. Further support for the default interventionist account can be found in the analysis of the responses (e.g., the heuristic scores) as there was no evidence of conflict sensitivity in the unlimited time condition. The no-conflict and conflict statements were answered in a similar time to one another, suggesting that the participants never noticed the difference between the two and employed similar strategies when answering them.

However, this is where support for the default interventionist account ends. When the experimental manipulation of severe time pressure was introduced the results showed significant evidence of conflict detection. Participants were more likely to respond heuristically in the absence of conflict, and logically when conflict was present in the statements. This provides support for the parallel-competitive model which states that people are able to detect conflict but are not always able to inhibit the fast heuristic answer (Sloman, 1996). The two systems of thought are believed to run in parallel to one another, thus alerting the reasoner when the they point towards different answers. The evidence from the limited time condition shows that people are sensitive to the conflict underlying heuristic and logical considerations and supports a widening body of evidence that conflict detection comes easily and naturally to people (De Neys, Cromheeke, & Osman, 2011; Bonner & Newell, 2010; De Neys, Moyens, & Vansteenwegen, 2010; De Neys et al., 2010; Villejoubert, 2009; De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2012; Ball, Phillips, Wade, & Qualyle, 2006; De Neys & Feremans, 2013). Overall, there was a high heuristic response to the task, which should not be attributed to a lax monitoring process. The high rate of heuristic responding was not due to reasoners not having acquired the relevant normative principles, or a failure to retrieve them (De Neys & Glumicic, 2008). If people were not taking the rule of conjunction probability into account, they would always respond heuristically to the conflict statements, which was not the case, as sometimes they responded logically. On some level, be it an active struggle or unconscious occurrence, it is clear that people under time pressure (i.e., despite being short of cognitive resources) are readily able to detect conflict, and this is evident due to their increased logicality when it comes to judging the conflict statements. If untrained participants are taking the rule of conjunction probability into account when making their judgements, and this is evident from their ability to detect conflict, then the problem is not lack of statistical sophistication. The problem is more likely the struggle to override the enticing heuristic, as seen in the work by Houde and Moutier (2003). In other words, they just cannot ignore the fact that Linda seems like a feminist. “People might not always manage to adhere to the norm, but they are clearly not simply discarding it or treating it as irrelevant…Clearly, people are more normative than their answers show.” (De Neys & Glumicic, 2008, p. 1280).

One point left to discuss is the evidence found that supports the newer logical interventionist account (De Neys, 2012). It is a more recent dual process theory that incorporates parallel processing to explain why people are so good at detecting conflict. However, instead of logical considerations emerging from slow, deliberate processing (e.g., traditional accounts of dual process theories); it proposes that people are able to arrive at a logical consideration by using intuitive reasoning. When put under severe time pressure to eliminate any differences in response times between the three judgement types, there was a significant increase in logical answers. In other words, participants relied less on heuristic considerations when making faster judgements. This finding lends support for the logical intuitionist account (De Neys, 2012; Villejoubert, 2009) as it shows that when people’s cognitive resources are limited by being put under severe time pressure (e.g., using System 1 reasoing), they tend to be more logical in their judgements than when they are given unlimited time and access to all their cognitive resources to deliberate (e.g., using System 2 reasoning). These findings support the idea that people can have an implicit and informal understanding of the law of conjunction probability (i.e., the first/fast response would also be the logical/correct one). The findings also showed that people were more likely to select the heuristic answer in the absence of time pressure, which could suggest that individuals who engage in effortful, deliberate thinking are also less prone to produce logical probability judgements, as if it was the slower, more deliberative form of thinking that was more vulnerable to prior beliefs and non-logical heuristics. To a point, these findings suggest that logical considerations can be fast and automatic while heuristic considerations can be slow and deliberate. The findings attempt to reconcile the apparent contradiction between the logical intuitionist account (De Neys, 2012, 2014; Villejoubert, 2009) and the parallel-competitive account (Sloman, 1996; Epstein, 1994), by suggesting that these two sources of logical responding might co-exist, depending on the time available to make the judgement and the cognitive resources available to the person making the judgement. Perhaps it is the logical judgement that comes first (i.e., an intuitive judgement) but when given more time to think this response is overridden by the compelling heuristic answer. Once the compelling heuristic has taken preference, logical responding requires more deliberative and effortful thinking to be reinstated. This speculation was supported by the fact that severe time pressure resulted in more logical responses (e.g., logical intuitions), whilst no time pressure resulted in higher representative answers (e.g., heuristic reflections). Additionally, the longer judgement latencies for logical responses to conflict statements in the unlimited time condition suggest that a huge amount of cognitive effort is required to reinstate the logical response after the persuasive heuristic response has taken hold. In a sense, putting participants under time pressure did not allow for the potent heuristic reflection to take hold of the reasoning process, as it removed the ability to reflect or deliberate. Thus, retaining the logical intuitive response. However, it is important to note the danger in assuming that a decrease in “heuristic” responding is a result of increased logical reasoning. The CCPJ task forced participants to make one of two responses, hence it is not clear that “logical” responding results from logical reasoning, or from a rejection of the heuristic response, or from participants simply giving up and guessing. These confounds in the data need to be addressed with future research to improve our understanding of the cognitive underpinnings of human judgement and decision making.

Despite there being some evidence that people high in analytical-rational style or low in intuitive-experiential style are less likely to commit reasoning errors and judgement fallacies (e.g., Pacini & Epstein, 1999; Shiloh, Salton, & Sharabi, 2002); the results of this study do not support this idea. Instead, they showed that people who rated themselves as medium on both the rational and experiential scales were actually less likely to commit the conjunction fallacy than those who rated themselves high on both scales. Also, it is important to note that this difference was only just significant (p = ,05). In fact, this data supports the findings of Lu (2015) who showed that rational and experiential cognitive styles did not influence the propensity for committing the conjunction fallacy. Some points to bear in mind when considering the REI is that it is a self-report measure, which brings certain problems along with it. Firstly, it relies on the honesty and introspective ability of the participants who might tailor their answers to fit a particular image, or perhaps they try to answer honestly but lack the introspective ability to provide an accurate response to a question. Participants may also vary according to their understanding or interpretation of a question. Additionally, there is the issue that intuitive thinking may not be accessible to conscious processing and; therefore, may be difficult to measure via self-report. It might be worth considering using a different task in future studies to determine which individual differences, if any, have an impact on people's logicality when making conjunction probability judgements. A task, for example, such as the Cognitive Reflection Test (CRT; Fredericks, 2005) which does not rely on self-report responses. Instead it measures through a series of three problems, people's ability to inhibit a heuristic intuitive response and replace it with a logical deliberate one.

Generalizabitily of Results. It is important to address a potential confound in the data. Kahneman (2003), Evans (2008) and Stanovich and West (2000), along with many other default-interventionist theorists, never argued that everyone would be biased in their judgements. In fact, most of the research states that approximately 80% of participants commit the conjunction fallacy, which leaves 20% of participants answering in accordance with logical assessments. Thus, all the default-interventionist theorists need to account for the currently lower heuristic scores in the conflict condition is a small proportion of reasoners who do not commit the conjunction fallacy. It is for this reason that “unbiased reasoners” were filtered out of the analysis. Filtering out people with perfect logicality (i.e., a zero heuristic score) may not control for this completely, but it should help. One assumes that people who systematically answer logically to all eight conflict statements are applying the objective rule of probability consciously and purposefully. They are using a qualitatively different type of judgment when making their answers and thus should be removed from the analysis. However, people who answer fewer than eight conflict statements logically show that they are paying attention to the scenarios and statements and using different types of processes to complete the task compared with those who are always logical. This highlights an important differentiation between those people who are logical and those people who provide logical answers. Figure 3. 6 shows the distribution of answers as a score out of eight ($N = 101$). There were no perfectly logical people in this sample, thus they cannot be held responsible for the lower heuristic scores in the conflict condition. Unfortunately, however, due to the design of the methodology one cannot guarantee that a few participants were not guessing. A decrease in heuristic responses does not simply mean that people are reasoning more logically. Conflict in the statements causes a situation of uncertainty, and it might be that under this uncertainty some people respond differently. They might be reasoning in line with logical assumptions, but they might also be rejecting the representative choice (i.e., they have no alternative other than to choose the logical answer), or they might be guessing.

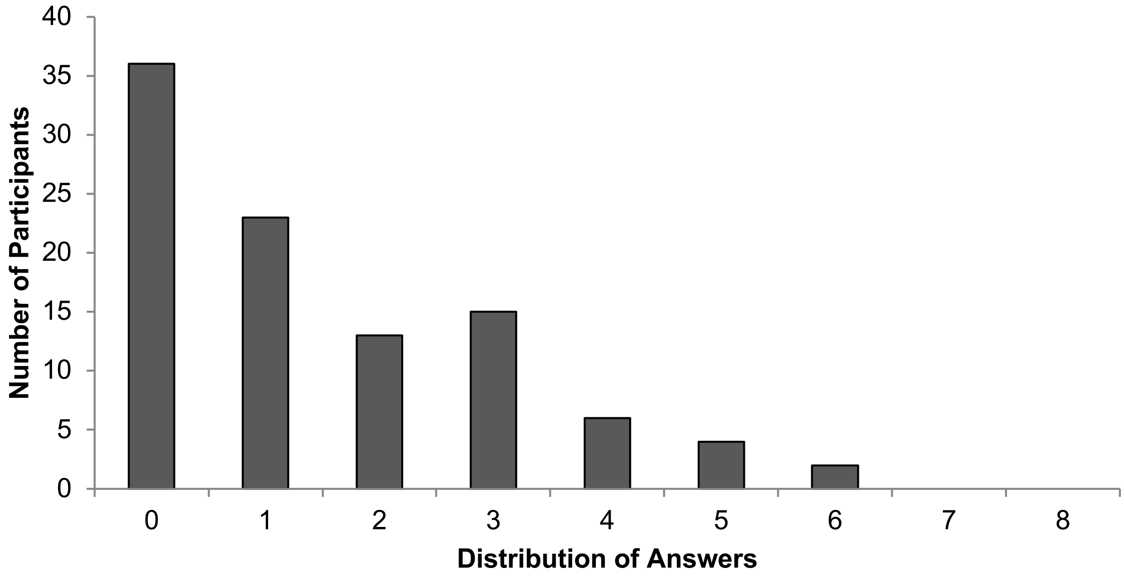


Figure 3. 6. Distribution of participant's logical judgements on the eight conflict statements (N = 101).

Final remarks and Future Directions. To conclude, the present research set out to test the controversial assumption that logical judgments emerge from intuitive and implicit processes whereas heuristic judgments arise from explicit reflective processes. The results did not fully support the hypotheses; however, they did shed new light on our understanding of the cognitive processes and the interplay between these processes when making social judgements. What we traditionally believe and know about the dual process systems and the ways in which they function might not be entirely correct. There is data that suggests System 1 might be able to produce both heuristic and logical responses. While the same can be said for System 2. However, these data are inconclusive and require further testing with a more robust methodology to eliminate any potential confounds. The results do however challenge the lax monitoring assumptions that people lack the ability, motivation, or resources to detect conflict between the two systems. In fact, there was solid evidence found for conflict sensitivity, lending support for a parallel model of reasoning as opposed to a serial one. The parallel-competitive model states that both logical and intuitive processes are activated from the beginning of a reasoning task, which leads to "simultaneous contradictory belief" when people suspect that something is wrong with their answer, but continue to make it regardless because they lack the ability, or motivation, to inhibit the pervasive intuitive heuristic response (Epstein, 1994; Sloman, 1996; Tversky & Kahneman, 1983). Further research might include a measure of confidence ratings following the CCPJ response to assess whether people do in fact experience feelings of rightness/wrongness with their answers.

In line with the logical intuitionist account (De Neys, 2012), the results suggest that people can intuit logical considerations, while it might be the more deliberative judgements that are susceptible to the representativeness heuristic (Handley & Trippas, 2015). However, these assumptions need further testing to rule out other explanations. One way of achieving this might be to redesign the CCPJ task. It was quite restrictive in the sense that it only offered two options to choose from. Participants were forced to choose one of those answers regardless of whether they agreed with them or not. A better option might be to include more choices in the test (see Travers, Rolison & Feeney, 2016). The authors presented the CRT test in a 4-option multiple choice format. For the conflict items, there was a correct logical option, the incorrect heuristic option, and two incorrect foil options. The no-conflict items presented the correct intuitive and logical option plus three incorrect foil options. A design such as this would assist with determining whether participants were in fact using logical or heuristic considerations when making their judgements, or whether they were simply confused and just making guesses, Future research might also further examine the factors, which may strengthen logical intuitions about conjunctive probabilities or lessen the impact of the heuristic processing on the individuating information. In line with the literature (Bago & De Neys, 2017; Pennycook & Thompson, 2012; Villejoubert, 2009), judgement latencies proved to be a successful method of encouraging intuitive thought while making judgements.

However, Evans, Dillon and Rand (2015) caution researchers not to interpret judgement latencies as a substitution for thinking styles. Thus, the following experiment incorporated an eye-tracker as a more fine-grained method of tracing systems of thought during conjunction judgements.

The importance of these findings leads to fostering better interventions to improve conjunction judgements. For example, biased social judgements may negatively impact a company's recruitment and selection procedures, as well as legal and medical decisions. The present findings call for further exploration of the idea that that stereotypical heuristics can be deliberately used to override more implicit, logical assessments of conjunctive probabilities.

Experiment 2: Eye-tracking Evidence for Logical Intuitions in Conjunction Probability Judgements

Using an adapted version of Villejoubert's (2009) methodology designed to disentangle logical and heuristic considerations, Experiment 1 showed that people who were put under severe time pressure to provide a judgement about a conjunctive probability statement were almost six times more likely to provide a logical answer. This finding suggests that fast intuitive answers are not necessarily biased while heuristic and biased answers are not necessarily fast. Still, when logical and heuristic considerations were in conflict, but participants were not pressured to make fast judgements, those who provided logical answers also took longer to respond. What is not clear is whether those longer response latencies were cued by heuristic considerations or logical considerations. Using response times to attempt to answer this question might not be sufficient. Evans et al. (2015) provide evidence through the use of social dilemma experiments that response times should not be interpreted as a direct proxy for the use of intuitive or reflective processes. While Gaissmaier, Ficic and Rieskamp (2015) explain that it is impossible to distinguish between serial and parallel processing using only a mean response time analysis. For example, one would expect shorter response times from parallel processed information compared to serially processed information. However, it is also possible that parallel models can increase in response time depending on the number of items being compared. Krajbich, Bartling, Hare and Fehr (2015) caution against using reverse inferences (e.g., the use of behavioural or biological measures to infer mental function) in research as it does not consider sources of variability in the data, such as favourable choice options.

Eye-tracking and judgement and decision making. Experiment 1 used judgement latencies and judgement responses as a proxy for measuring conflict detection, but this is an indirect inference; a more direct test would require tracing information acquisition. Thus, instead of using response times one could explore other process-tracing methods such as eye-tracking. Being able to retrieve information from our environment and memory is crucial for every decision we make. Our eyes are the foremost organs for acquiring information from our environment, and this acquisition is achieved when the eye fixates an object (Russo, 2011). Tracking the eye's movements during decision making yields a fixation sequence, which provides a detailed trace of the information used. There are many advantages for using eye tracking methodology for tracing processes. Eye movements do not require much effort at all, in fact, they are so "metabolically cheap" that they are sometimes used to substitute working memory as a storage retrieval system (Spivey, Richardson & Dale, 2009, p. 237; Droll & Hayhoe, 2007). Eye fixations can be recorded unobtrusively and are difficult to censor due to their speed and automaticity. There are however also disadvantages including highly sophisticated and expensive equipment and the task of interpreting the data. Eye fixations tell researchers where participants are looking, not what they are thinking; which is tricky when one ultimately needs to know the latter to develop and test observed phenomena and theories (Russo, 2011). Perhaps the most common use of fixation locations is interpreting their frequency as an index of importance. In other words, the more a person fixates an object, the greater that object's importance (van Raaij, 1977; Reisen, Hoffrage & Mast, 2008; Gidlof, Wallin, Dewhurst & Holmqvist, 2013). While the average fixation duration is commonly interpreted as an indicator of the processing depth or effort, with longer latencies assumed to indicate more complex cognitive processes than shorter ones (Ableson & Levi, 1985). Pieters and Warlop (1999) found that fixation durations on the selected item were always longer than the not selected alternatives, even under time pressure. Horstmann, Ahlgrimm and Glöckner (2009) used eye-tracking technology as a fine-grained method to analyse how the instruction to decide intuitively or deliberately affects information search and integration. They used fixation durations in an attempt to identify the difference between intuition and deliberation; however, they found no difference in durations between the two, leading to the assumption that intuition and deliberation share similar underlying process modes.

Previous gaze and eye-tracking studies showed that longer latencies are associated with longer inspection of normatively critical problem information in syllogism tasks (Ball, Philips, Wade & Quayle, 2006). These researchers analysed eye-movements while participants were solving deductive syllogisms. These types of problems can be constructed to contain conflict, or no conflict. Conflict is present when the logical validity of an argument structure conflicts with the believability of its conclusion (e.g., All mammals can walk. Whales are mammals. Thus, whales can walk). When conflict is detected, it is believed that the analytic System 2 must intervene. The results of the eye-tracking study showed that people spent longer inspecting the syllogistic problems containing conflict than those without conflict. The source of conflict in conjunction fallacy tasks remains to be identified.

The conjunction fallacy and working memory. Experiment 1 showed limited evidence that self-reported individual differences in thinking processes were related to conflict detection. Another approach would be to examine more objective individual differences such as working memory capacity. **“**The term working memory refers to a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning” (Baddeley, 1992, p. 556). A complex working memory task measures the number of items that can be recalled. These tasks have simultaneous storage and processing components: participants are expected to maintain information in an active state for later recall, whilst manipulating information for a current computation (Barrett, Tugade, & Engle, 2004). There are many models of working memory in existence; however, the one most often identified with is that of Baddeley (1992, 2007) who proposed three components: the central executive, the phonological loop and the visuospatial sketchpad. The central executive acts as a control centre directing relevant information or suppressing irrelevant information from the phonological loop and visuospatial sketchpad. The phonological loop stores the sounds of language and prevents the information form decaying by continuously refreshing it in a rehearsal loop. For example, remembering a telephone number by constantly repeating it to yourself. The visuospatial sketchpad stores visual and spatial information, such as shapes, colours, textures and locations; all of which can be manipulated to construct mental maps. These two constructs – the phonological loop and the visuospatial sketchpad – can work simultaneously without affecting the ability of the other (Denis, Logie & Cornoldo, 2012). More recently, Baddeley (2000) added a fourth component to the model of working memory: the episodic buffer. It is believed to be a limited-capacity temporary storage system that integrates information from various sources. This passive system is dedicated to linking information across domains to form cohesive units of information (i.e., visual, spatial and verbal) with time sequencing, and it "serves as an interface between a range of systems, each involving a different set of codes" (Baddeley, 2000).

As mentioned earlier, theorising in the heuristics and biases literature has emphasised dual-process models of cognition, which embody the assumption that there are two thinking systems: System 1, which operates intuitively and autonomously and System 2, which operates deliberatively and non-autonomously (Kahneman & Frederick, 2002). The presence of thinking biases is extremely prevalent and common place; however, the ability to suppress an erroneous response and override it with a correct one varies from individual to individual. Evans and Stanovich (2013) believe that differences in working memory capacity and intelligence influence resistance to belief biases. They highlight how studies of individual differences in reasoning have shown for heuristics and biases tasks, that "the modal response displays negative correlations with cognitive sophistication" (p. 234). The common assumption is that a person requires a certain amount of cognitive/computational power to override a heuristically primed response, and this ability to recognise the need to override (or detect the conflict between heuristic and logical considerations) might be related to intelligence. This would result in a negative relationship between biased responding and cognitive ability (Stanovich and West, 2008). However, these researchers also found that people with high SAT scores (an indicator of cognitive ability) were more susceptible to the conjunction fallacy. One can conceive that after detecting the conflict between an intuitive and logical assessment, people with strong cognitive abilities will have no problem in engaging deliberative thinking, and it is from this deliberation that they are arriving at their heuristic answers. This evidence suggests the possibility that the heuristic response comes from the deliberative system. While Chen and Sun (2011) found that people with reduced working memory capacity were more likely to use a satisficing heuristic when making financial decisions. However, this may not apply to the conjunction fallacy as results from Experiment 1 suggest that people make fewer errors when relying on intuitive System 1 thinking (i.e., when they are put under severe time pressure). This implies that people can implicitly learn the logical rule through experience but overcome it when they start thinking about the Linda task. Weldon, Mushlin, Kim and Sohn (2013) point out that working memory capacity has been consistently correlated with performance in a variety of cognitive tasks. Such as higher working memory capacity reflects better ability to focus attention, filter out distractions, and maintain task goals, which according to the authors, implies that working memory capacity "may directly reflect the ability to exert cognitive control" (p. 6). Similarly, according to Stanovich (1998, see also, Evans & Over, 1996), the ability to resist belief-based responding is highly dependent on the cognitive resources of the explicit system, such as working memory. While Handley, Beveridge, Dennis and Evans (2004) found that children with higher working memory respond more rapidly and with greater accuracy to reasoning tasks. If the classic account of the conjunction fallacy is true (i.e., the logical answer requires inhibiting the heuristic answer); and inhibition is a feature of working memory; so higher working memory capacity should be associated with *less* heuristic responding. On the other hand, the alternative account suggested by the data in Experiment 1 (i.e., less heuristic responding under time pressure, assuming time pressure means less opportunity for working memory-intensive deliberate processing), predicts that higher working memory capacity should be associated with *more* heuristic responding.

The Present Experiment. On the one hand, it is traditionally believed that time pressure promotes intuitive thinking (Capraro & Cococcioni, 2016; Sloman, 1996; Kahneman, 2011; Evans & Stanovich, 2013; Stanovich & West, 1998), and Experiment 1 found that people were more likely to provide a logical answer under time pressure, which lends support for the logical intuitions account, in other words, the logical answer comes first. However, on the other hand recent studies have warned against using time pressure to induce intuitive thinking because response speed interacts with decision conflict (Evans, Dillon & Rand, 2015), strength of preferences (Krajbich, Bartling, Hare & Fehr, 2015), and peoples' probability to err when implementing a strategy (Recalde, Riedl & Vesterlunde, 2014). All of which could be problematic when exploring judgement under time pressure. However, the latter finding (i.e., the likelihood of making a wrong choice under time pressure) was not supported by Experiment 1, which showed that time pressure increased logical responses. A criticism of Experiment 1 is the fact that it only measured conflict through behaviours and response time (i.e., the reduced proportion of heuristic answers). However, there is evidence as mentioned earlier that people's behaviour may not be a good measure of conflict detection. Thus, the experiment reported here used a more fined-grained measure of conflict processing than response latencies and time pressure, namely eye movements. Eye-tracking methodology can be used for tracing intuitive and deliberative processes (Glöckner &

Herbold, 2011; Rayner, 1998). Specifically, it is assumed that automatic, superficial levels of processing have shorter fixation latencies, while deeper processing and a deliberate consideration of information is related with longer fixations (Glöckner & Herbold, 2011; Velichkovsky, Rothert, Kopf, Dornhofer & Joos, 2002). Furthermore, eye tracking permits the detection of conflict in processing as eye movements are known to be disrupted when readers' general knowledge conflicts with the text they are given to read (e.g., Duffy & Keir, 2004; Featherstone & Sturt, 2010).

Based on this information and in the hopes of shedding light on the processes involved in judging conjunctive probabilities, the aim of this experiment was to uncover whether individual differences in patterns of eye movements could account for differences in individuals' propensity to generate logical probability judgements and focussed in particular on the question of conflict detection between heuristic and logical considerations. If the assumption is true that higher rates of fixations and longer durations of fixations equals higher importance/attention, then one could anticipate that people will fixate more and for longer durations on stimulus that they deem to be important when retrieving their answers. The main hypothesis of this experiment was therefore that eye movements would differ under conflict and non-conflict judgment settings. Thus, this study aimed to uncover whether individual differences in patterns of fixations for the different types of probability statements could account for differences in individuals' propensity to generate a logical judgement.

The question is whether conflict is only detected by people who have the cognitive means to do so (i.e., a high intelligence and strong working memory). If according to the default-interventionist view committing the conjunction fallacy is largely unconscious and in fact due to a lack of cognitive ability, then one would expect measures of effortful cognitive processing such as working memory capacity to be largely irrelevant in the process of solving conjunction judgements. However, if processing is based on a conscious analysis of the constituent events of the problem and their relation to the description vignette, then one would expect working memory to be positively correlated with successful judgement performance. This research study is exploring the idea that intuition can reflect the laws of logic, in other words the logical answer comes first when making judgements, while deliberation might be more susceptible to heuristic biases. It assumes that people are readily able to detect the conflict underlying logical and heuristic considerations. If this assumption is indeed the case, then one could expect to see similar results to Chen and Sun (2011). To be more specific, people with lower working memory capacity will rely more heavily on intuitive processes when making their conjunction judgements, thus participants who exhibit lower working memory capacity will also exhibit higher logicality in their conjunction judgements. Stanovich and West (2008) found, people with higher SAT scores also committed the conjunction fallacy more regularly. If the classic account of the conjunction fallacy is true (i.e., the logical answer requires inhibiting the heuristic answer), higher working memory capacity should be associated with less heuristic responding. On the other hand, the alternative account (suggested by the data in Experiment 1), predicts that higher working memory capacity should be associated with more heuristic responding. Thus, participants who exhibited lower working memory capacity were expected to also exhibit higher logicality in their probability judgements.

Method

Participants. Forty-three psychology students from Kingston University were recruited for this study through the Sona-System – a Cloud-based Participant Management Software. Two students were already familiar with the Linda problem and therefore were removed from subsequent analysis. Of the remaining 41 participants (6 men and 35 women; mean age = 27.93, *SD* = 3.48), 29% were postgraduates and 71% were undergraduates. All the participants had background knowledge in psychology. The majority of the participants were enrolled only in psychology modules (66%); however, some had dual majors which included criminology (7%), business (7%), human biology (5%), sociology (5%), journalism (2%), human rights (2%), English (2%), and creative writing (2%). The experiment was conducted in English, 61% of the participants were native English speakers. Participation was voluntary and anonymous, and the study took approximately 45 minutes to complete. Participants were either paid £6 or received course credits for their time and participation.

Apparatus and Materials. Materials were adapted from those used in Experiment 1. Sixteen CCPJ tasks were programmed in Matlab and presented to each participant on a 21-inch CRT display screen. Participants' eye movements were recorded using a head mounted video-based Eyelink II eye tracker with a spatial resolution of 0.5 degrees and temporal resolution of 2ms while they were assessing the validity of the statements presented.

Each CCPJ trial consisted of a thumbnail description and a statement sentence. The descriptions were short paragraphs that outlined the personality traits, hobbies and vocations of fictitious people according to recognised stereotypes (e.g., a feminist is single, outspoken, studied Philosophy at university and is deeply concerned with issues of discrimination and social justice). Each description was followed by a unique statement; one sentence that contained both a single clause and a conjunction clause. Within the single and conjunction clauses there was always a word that aligned with the stereotype (e.g., feminist) and a surprising word that contradicted the stereotype (e.g., bank cashier). Participants were required to either accept the statement as correct, or reject it as incorrect depending on whether they believed the single clause, or the conjunction clause, to be more likely to occur. The format of the statements could be manipulated by the arrangement of both the stereotype and surprise words within the single and conjunction clauses. There were four statement categories: Representative and Logical (R&L), Representative but Illogical (R/I), Unrepresentative but Logical (U/L), and Unrepresentative and Illogical (U&I; see Table 2. 1, p. 74 for an illustration of the four types of statements for the CCPJ task).

Manipulation checks were conducted on the CCPJ tasks to ensure that the 16 problems were similar to one another. Thus, any differences found in reading latencies or eye-movements could be assumed the result of experimental manipulations and not lack of control in the experimental design. The number of words was counted for each thumbnail description and for each statement sentence, and an average word count was computed for each description and statement type (i.e., one value for each of: R&L, R/I, U/L, U&I). The means for the descriptions and statements were each subjected to a repeated-measures ANOVA with four levels (statement type). Results confirmed there was no significant difference in word count between the 4 types of description vignettes ($M_{R\&L} = 37.50$, $SD_{R\&L} = 8.23$, $M_{R/I} = 34.75$, $SD_{R/I} = 5.32$, $M_{U/L} = 38.75$, $SD_{U/L} = 5.62$, $M_{U\&I} = 39.50$, $SD_{U\&I} = 2.38$), $F < 1$; nor the 16 statements ($M_{R\&L} = 18.75$, $SD_{R\&L} = .96$, $M_{R/I} = 19.50$, $SD_{R/I} = 1.29$, $M_{U/L} = 19.00$, $SD_{U/L} = .82$, $M_{U\&I} = 19.00$, $SD_{U\&I} = 1.41$), $F < 1$. Accordingly, we can assume that any differences observed in the data are the result of individual differences in behaviour. The descriptions and statements remained unchanged between Experiment 1 and 2. However, some of the longer names of the fictitious characters were modified or simplified into shorter one syllable names. For example, Linda became Lynn and Samantha became Sam. This was done to ensure eye movements were as stable across the sixteen scenarios as possible.

Design and procedure. This study used a 2 (Areas of Interest: Description AoI vs. Statement AoI) x 2 (Conflict: Present vs. Absent) repeated measures design. It examined the effect of conflict on the eye movement patterns across the following areas of interest (AoIs): description vs. statement and single clause vs. conjunction clause. The experiment took place in a psychology laboratory on the Kingston University campus. Upon giving their informed consent, all participants completed both the CCPJ problem and the computation span.

CCPJ task. Participants were sat in front of a 21-inch CRT display screen while the experimenter set up the Eyelink II. The eye-tracker was placed and secured on the head of the participants, and the video cameras were positioned to get a clear and accurate reading of the pupils. The experimenter then performed the calibration and instructed the participants to keep their heads as still as possible for the remainder of the CCPJ task. The participants' heads were rested on a chin rest to ensure as little movement as possible. The first slide of the CCPJ task presented the instructions to the participants: "In this experiment, you will be presented with a series of scenarios describing a person followed by a statement. Your task is to read these descriptions and then indicate if you think the accompanying statement is true or false". At this point the experimenter explained that if they wanted to answer true, and accept the statement, they must click on the left mouse button, and if their answer was false, and they wanted to reject the statement, they must click on the right mouse button. The next screen explained to the participants they would be given three examples trials before beginning the experiment. This allowed them to become familiar with the task and understand what was expected of them before beginning the experimental problems. They were prompted to take as much as they needed to read the descriptions, and to consider the information carefully before reaching a decision. After completing three examples, they completed the 16 CCPJ problems in a random order while their behavioural responses and eye movements were recorded. More specifically, their responses and time latencies for each CCPJ problem, as well as percentage dwell time, total dwell time, first run dwell time, refixation number and refixation dwell time for six areas of interest (AoI). The first AoI was the description vignette, and the second was the statement sentence. The statement AoI was segmented into two additional AoIs, which allowed us to examine dwell times and refixations for the single and conjunction clauses. We safeguarded that no AoIs within the statement sentence were broken up or fragmented by falling over two lines. For example, you will see in Figure 3. 7 that the conjunction clause, “cashier and feminist”, falls on the same line. The 3 example trials and the 16 CCPJ problems were all presented in the same format.

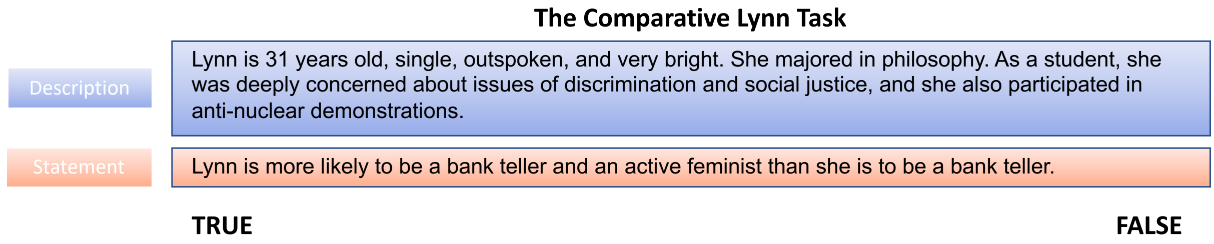


Figure 3. 7. Design and layout of a Comparative Conjunction Probability Judgment problem. The blue area defines the description Area of Interest (AoI), while the orange area defines the statement AoI. *Note*. The participants did not see the AoIs when they completed the task; they simply read the descriptions and statements in black text on a white background

Computation span. This task was adapted from Ashcraft and Kirk (2001) and required participants to solve a series of arithmetic problems while also remembering the last digit from each problem. For example, participants were presented with simple arithmetic problems: “5 + 2 = ?, 9 – 6 = ?”, for which they were required to give the answers: “7, 3”. Following this they had to recall the last number of each problem for each of the several problems in that trial, in order: “2, 6”. The amount of problems per trial increased with the advancement of each trial. Thus, the span task required both on-line processing for arithmetic problem solutions, simultaneously combined with storage and maintenance of information in working memory for serial recall (Ashcraft & Kirk, 2001). The task was presented in video format using Windows Media Player software. Participants received instructions to the task both visually and verbally. They read the instructions off the computer screen whilst simultaneously listening to an audio recording (i.e., a voice over) of the same words incorporated into the video. This was done to ensure that all participants received the same introduction to the task, and also ensured they paid attention to the instructions and knew what was expected of them during the task. Participants were given two practice trails before the task began to ensure they felt familiar and comfortable with the task. The verbal voice recording stopped after the instructions and practice trials; it was not present during the task trials.

Participants were instructed to read the simple arithmetic expression and told to announce their answers aloud. They were also asked to remember the second number for each equation presented as they would be required to recall these target digits after each series of arithmetic problems. The arithmetic problems were presented one at a time on the computer screen. Participants read the problems, processed the answers mentally and expressed their answers verbally. After each series of arithmetic problems, a blank page was displayed, and participants were asked to verbally recall the target digits in the exact order in which they were presented. Time allowed for recall was 4 seconds per target digit. Instructions emphasised the importance of answering the arithmetic problems correctly as accurate recall would only be recorded if they had given the correct answer to the arithmetic problem. The amount of arithmetic problems presented on each trial increased successively from one to seven, and participants completed two trials of each series length. In line with Salthouse and Babcock (1991), the arithmetic problems were all addition and subtraction equations of two digits between 1 and 9 each. Further restrictions prescribed that the answers to the problems could not equal negative amounts, the second digit could not be the same for two adjacent problems in a trial, and the answer to the problem could not equal the amount of the second digit. Participants had to process each equation as well as recall the relevant digit correctly to score a point.

Results

Outliers. A total of 41 Kingston University students participated in the experiment (6 men and 35 women; mean age = 27.93, *SD* = 3.48). The sample was screened for participants who were always logical (i.e., always chose the logical answer), as well as participants who always answered incongruently (e.g., rejected statements that were both representative and logical, instead of accepting them) on non-conflict trials. No reasoners were always logical, nor gave judgement responses that were always incongruent, thus the sample size remained at $N = 41$.

Following the procedures outlined by Meade and Craig (2012), an overall "flag score" was computed to identify careless responders who were identified as outliers on three different criteria. First, a total duration score was computed by adding the total description dwell time to the total statement dwell time, then dividing by 1000 to create a score measured in seconds ($M = 219.65$ seconds, $SD = 61.10$, min = 103.96, max = 376.86). Total duration in seconds was then converted into a z-score. Outliers were identified as any cases with z-scores outside the (-3,3) range. No outliers were identified; thus, no flags were allocated for total duration. Secondly, participants were screened for multivariate outliers in their reading times of the two areas of interest: descriptions and statements. Mahalanobis scores showed that one participant was flagged as an outlier for their reading time of the descriptions, but there were no outliers flagged for statement reading times. In line with Meade and Craig (2012), a total flag score was computed for the participants (total duration in seconds flags + description AoI flag + statement AoI flag). There were no participants who achieved a total flag score of 2 or higher, thus the sample remained at 41 participants.

Behavioural analysis

Conflict sensitivity. To examine conflict sensitivity, a heuristic score was computed for each trial to assess whether the participants were sensitive to the conflict between heuristic considerations (i.e., representativeness) and logical considerations when making their judgements. As in Experiment 1, responses were coded as heuristic when participants accepted representative statements (R&L, R/I, see Table 2. 1, p. 74) or rejected unrepresentative statements (U&I, U/L). A total heuristic score was computed for both the statements containing conflict and those without conflict. Heuristic scores ranged from 0% (never followed heuristic assessment) to 100% (always followed heuristic assessment) and were analysed using a paired samples t-test. Results showed that although participants answered more logically when conflict was present in the statements compared to when conflict was absent, this difference was not significant, $M_{\_noconflict} = 86.59$, $SD = 16.39$, 95% CI [81.41, 91.76], $M_{\_conflict} = 83.23$, $SD = 22.12$, 95% CI [76.25, 90.21], $t(40) = 0.96$, $p = .343$, Cohen's $d = 0.17$. Thus, it was not possible to conclude that participants were sensitive to conflict between heuristic and logical considerations on the sole basis of the behavioural data in this experiment.

Judgement latencies. Response latencies were analysed to assess which responses showed sensitivity to conflict between heuristic and logical considerations. Assuming that people who always respond with a heuristic judgement under conflict use a different strategy overall, we first screened and excluded them so we could compare latencies for heuristic and logical answers under conflict ($n = 19$, 46%). In line with previous research (Pennycook, Fugelsang, & Koehler, 2012; De Neys & Glumicic, 2008) and in accordance with Experiment 1, three average response latency scores were computed for each participant: non-conflict (responses to R&L and U&I statements), conflict-heuristic (heuristic responses to R/I and U/L statements) and conflict-logical (logical responses to R/I and U/L statements). Latencies were converted to $\log^{10}$ prior to analysis to normalize the distribution and subjected to a

3(judgment type: non-conflict vs. conflict-heuristic vs. conflict-logical) repeated measures ANOVA. The results revealed no difference in latencies across judgement types, $M_{\_noconflict}$ = 12.31 seconds, *SD* = 3.35, 95% CI [10.79, 13.83], $M_{\_conflict\_heuristic}$ = 12.05 seconds, *SD* = 3.31, 95% CI [10.55, 13.56], $M_{\_\mathrm{conflict\_logic}}$ = 12.57, *SD* = 3.76, 95% CI [10.85, 14.28], $F(2,40)$ = 0.321, p = .727, η_p^2 = .02. Once again, we found no evidence for conflict sensitivity between heuristic and logical considerations on the sole basis of the behavioural data in this study (i.e., responses and response times).

Although no main effect was found for conflict sensitivity using the ANOVA calculation, the data in fact lends itself better to a correlation as both independent and dependent variables are continuous. Thus, a Pearson's correlation was computed to assess the relationship between task duration and logicality when judging conjunction probability. A logical score was computed for each participant, which represented the amount of times they answered in line with logical considerations when answering the statements containing conflict. It ranged from 0 (never logical) – 8 (always logical). A weak, negative and marginally significant relationship was found between these two variables, r = -0.28, n = 41, p = .079. A scatterplot summarises the results (see Figure 3. 8). There appears to be a trend emerging in the data that the faster participants performed on the task, the higher their logicality scores were. This trend lends support for the idea that the fast judgements result in responses aligned with logical considerations, and not the slower judgements as previously assumed.

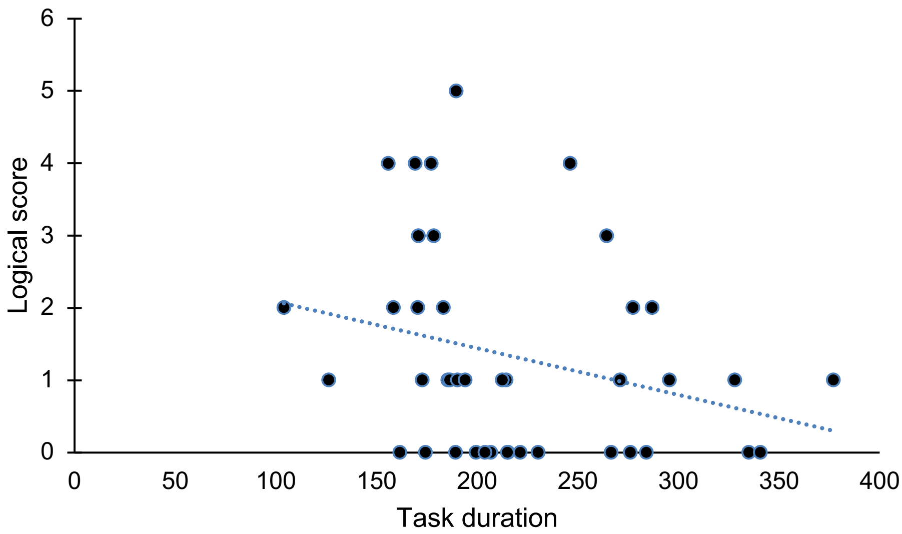


Figure 3. 8. The relationship between mean duration latencies (in seconds) and logicality on conflict statements (score out of a possible 8). $N = 41$

Individual differences

Computation Span Task. The logical score was correlated against the participants' computation span scores which ranged from 0 (zero recall) – 56 (perfect recall). A Pearson's correlation was computed to assess the relationship between computation span and logicality when judging conjunction probability. Surprisingly, no relationship was found between these two variables, $r = 0.21$, $n = 41$, $p = .196$. A scatterplot summarises the results (see Figure 3. 9). Overall, a strong or weak working memory could not predict logicality on the task. Next, I analysed the eye movement data.

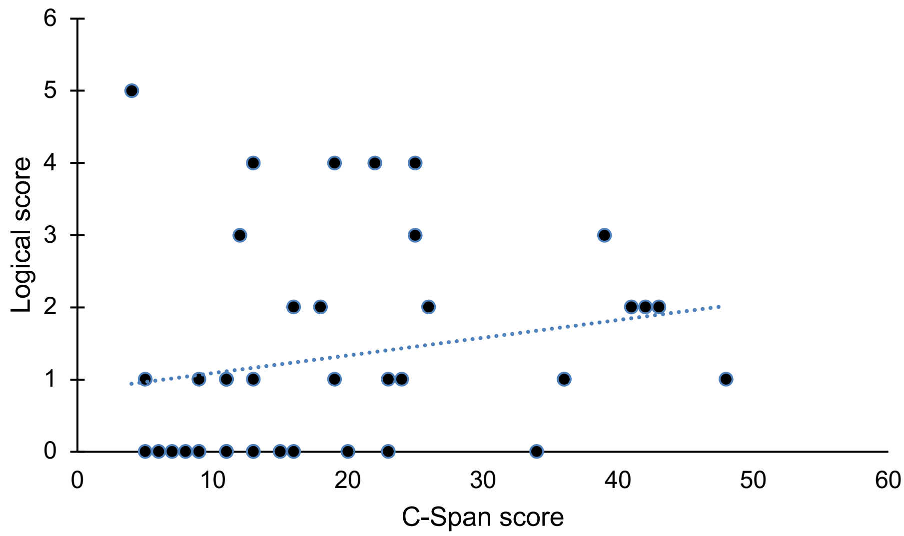


Figure 3. 9. Relationship between logicality (logical score) versus working memory (C-span score). $N = 41$

Eye-tracking analysis

Eye movement patterns were examined to assess whether participants were sensitive to the conflict between heuristic considerations and logical considerations within the CCPJ task. The number of refixations (i.e., number of times participants revisited an AoI with at least one fixation outside that AoI) and the refixation dwell times (i.e., total amount of refixation time within an AoI) were examined across both of the conflict conditions (conflict absent vs. conflict present) and the AoIs (description vs. statement, and single clause vs. conjunction clause).

Number of refixations. Participants redirected their eyes more often towards the description in the absence of conflict; however, they redirected their eyes more often to the statement in the presence of conflict (see Table 3. 2 below for descriptive statistics).

Table 3. 2

Means, Standard Deviations and 95% Confidence Intervals of number of refixations as a function of the AoIs and conflict conditions

		M	*SD*	95% CI	
				Lower Bound	Upper Bound
AoI	Description	3.57	1.29	3.16	3.97
	Statement	3.67	1.60	3.16	4.18
Conflict	Absent	3.46	1.23	3.07	3.85
	Present	3.78	1.37	3.35	4.21

A 2 (AoI: description vs. statement) x 2 (Conflict: absent vs. present) repeated measures ANOVA confirmed the overall difference in number of refixations between AoI's was not statistically significant, $F(1,40) = 0.18$, $p = .673$, $\eta_p^2 = .01$. There was however, a significant main effect of conflict condition, $F(1,40) = 6.13$, $p = .018$, $\eta_p^2 = .13$, showing that participants refixated more often when there was conflict present in the trials compared to when conflict was absent. There was also a significant interaction between AoI and conflict, $F(1,40) = 34.40$, $p < .001$, $\eta_p^2 = .46$. Participants refixated more often on the description when conflict was absent, but more often on the statement when conflict was present (see Figure 3. 10). Paired samples t-tests confirmed that participants refixated more often on the description when conflict was absent, $t(40) = 3.01$, $p = .004$, Cohen's $d = .24$, and conversely on the statement when conflict was present, $t(40) = -4.49$, $p < .001$, Cohen's $d = .55$.

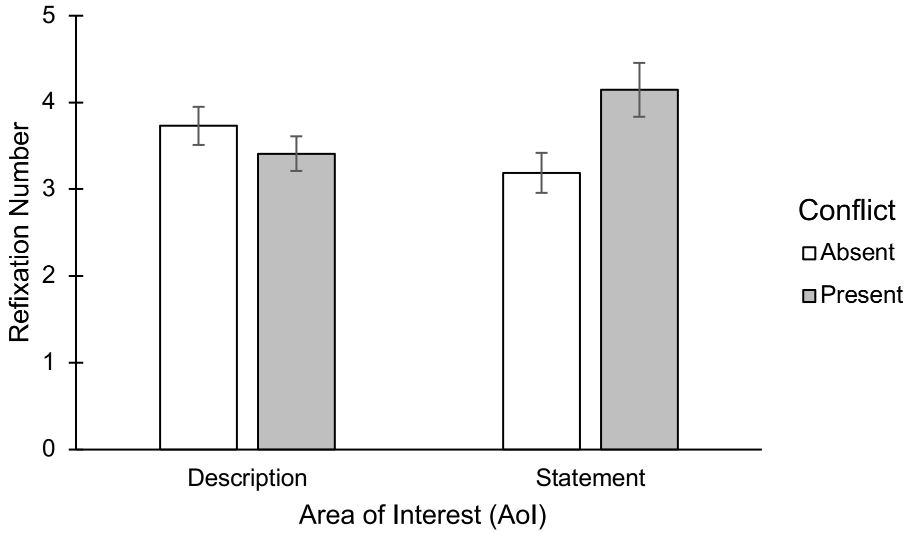


Figure 3. 10. Mean number of refixations as a function of AoI (description vs. statement) and conflict (absent vs. present). *Note*. Error bars represent standard error of the mean

When examining eye refixations and refixation dwell times it might be possible to notice patterns in people's eye movements across the four different statement types. For example, in R&L statements, the representative single event (feminist) is more likely and more representative than the conjunction (bank teller and feminist) so we might expect less refixations on the conjunctive event (because it is easy to dismiss). By contrast, in R/I statements, the conjunction (bank teller and feminist) is more representative than the unrepresentative single event (bank teller), but bank teller is more probable. In this scenario, one should expect more refixations on the conjunctive event (because it is harder to dismiss). Similar findings would be expected from the U/L and U&I statements; however, this might not be the case as it might be harder to interpret unrepresentative events. For example, in U&I statements the conjunction is less likely and less representative than the single event, but the sentence says it is not, so the conflict in this statement might be the qualifier "more likely". Thus, an exploratory examination of the patterns of eye movements across the different statement types was carried out. No inferential statistics were conducted because I did not have a priori hypothesis. This was purely an exploratory exercise. The graphs below show no specific patterns across the four types of statements; however, they do show that participants refixated their eyes more often towards the conjunction clause in general, and they also refixated more often when conflict was present in the statements (see Figure 3. 11). Participants also refixated more on the conjunction clause in general, as well as the statements containing conflict compared to those without conflict (see Table 3. 3).

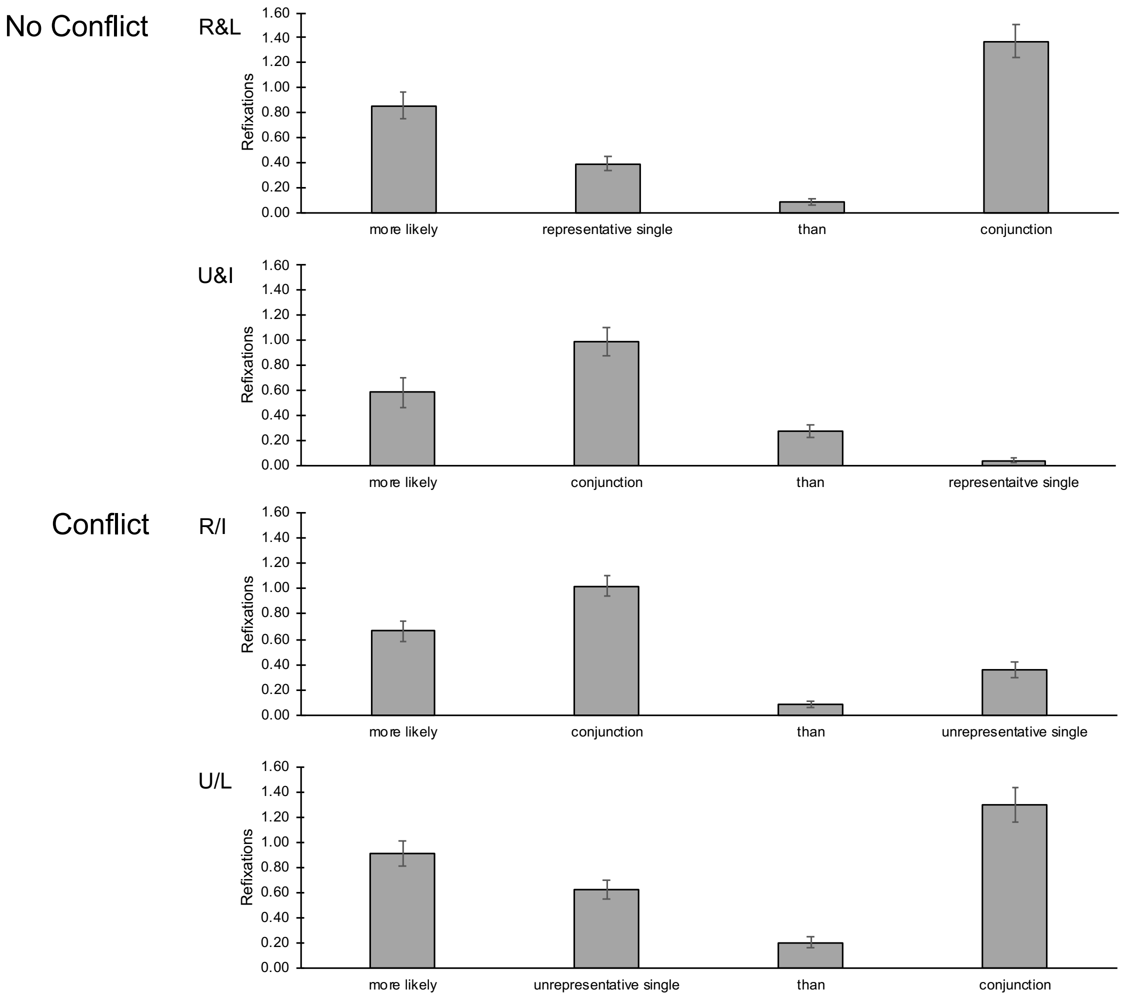


Figure 3. 11. A descriptive analysis highlighting the number of refixations for each sentence component across each of the statement types

Table 3. 3

Means, Standard Deviations and 95% Confidence Intervals of number of refixations for the AoIs and conflict conditions

		M	*SD*	95% CI	
				Lower Bound	Upper Bound
AoI	Conjunction clause	1.19	0.51	1.03	1.35
	Single clause	0.35	0.22	0.28	0.42
Conflict	Absent	0.62	0.31	0.52	0.72
	Present	0.92	0.45	0.78	1.06

Dwell times. In addition to refixating more often as outlined above, participants also gazed for longer on the description in the absence of conflict; however, they gazed for longer on the statement in the presence of conflict (see Table 3. 4 below for descriptive statistics).

Table 3. 4

Means, Standard Deviations and 95% Confidence Intervals of refixation dwell times for the AoIs and conflict conditions. Note. Latencies are expressed in milliseconds

		M	*SD*	95% CI	
				Lower Bound	Upper Bound
AoI	Description	4349.85	1799.21	3781.95	4917.76
	Statement	2182.60	1032.74	1856.63	2508.57
Conflict	Absent	3274.25	1415.51	2827.46	3721.04
	Present	3258.21	1244.79	2865.30	3651.11

A 2 (AoI: description vs. statement) x 2 (Conflict: absent vs. present) repeated measures ANOVA confirmed a significant difference in refixation dwell time between AoI's, $F(1,40) = 97.90$, $p < .001$, $\eta_p^2 = .71$. Participants' refixations lingered for longer periods of time on the descriptions versus the statements. There was no overall significant main effect of conflict condition, $F(1,40) = 0.02$, $p = .881$, $\eta_p^2 = .001$. However, there was a significant interaction between AoI and conflict, $F(1,40) = 21.98$, $p < .001$, $\eta_p^2 = .36$. Participants refixation dwell times were longer when looking at the description when conflict was absent, but when conflict was present their refixation dwell times were longer on the statements (see Figure 3. 12). Paired samples t-tests confirmed that participants refixated more often on the description when conflict was absent, $t(40) = 3.12$, $p = .003$, Cohen's $d = .25$, and conversely on the statement when conflict was present, $t(40) = -3.19$, $p = .003$, Cohen's $d = .39$.

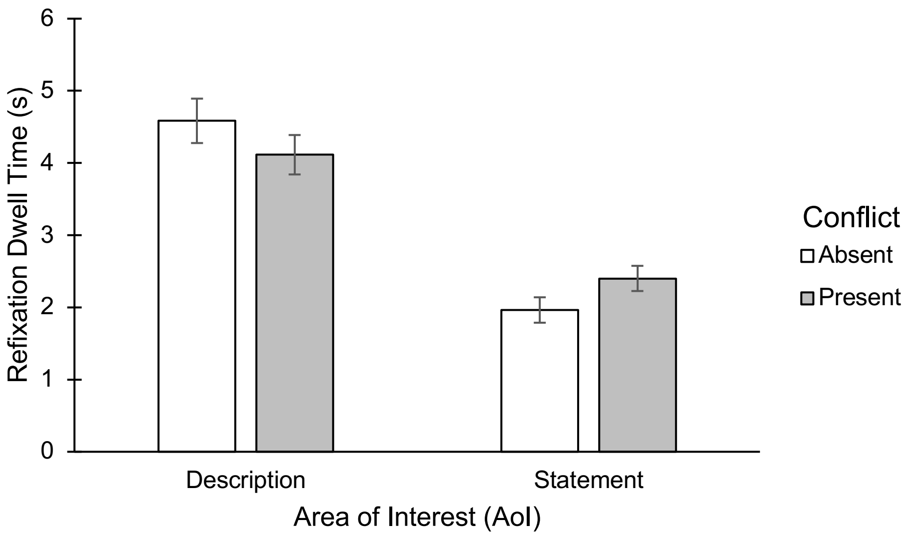


Figure 3. 12. Mean refixation dwell times as a function of AoI (description vs. statement) and conflict (absent vs. present). *Note*. Error bars represent standard error of the mean

A similar exploratory exercise to the one above was performed on the refixation dwell times to see whether any patterns in the eye movements latencies could be found across the four types of statements. Participants seemed to dwell for longer periods of time on the conjunctions in general, but no other stand out patterns were found. Conflict did not seem to make a difference to the refixation dwell times. See Figure 3. 13.

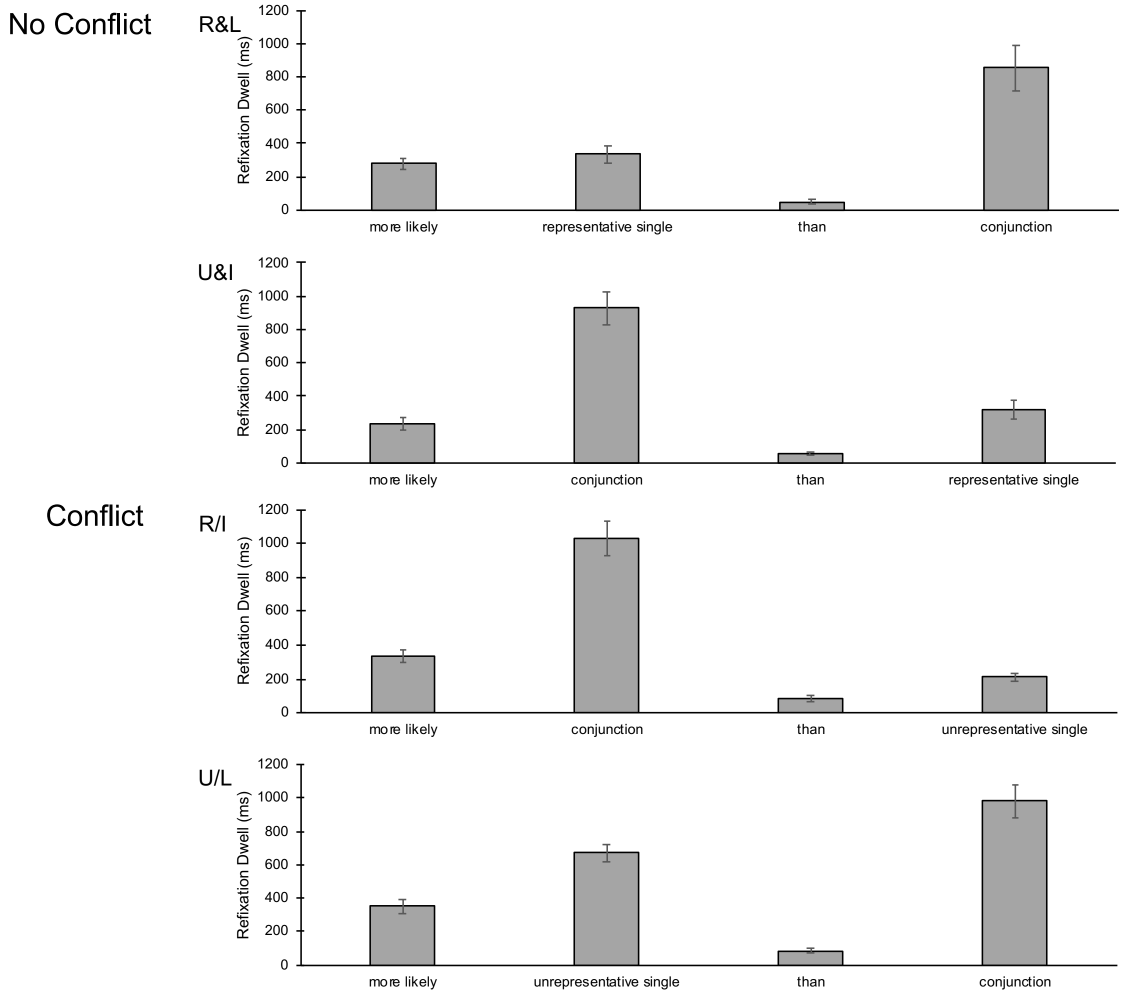


Figure 3. 13. A descriptive analysis highlighting the number of refixation dwell times for each sentence component across each of the statement types. *Note*. Dwell times are represented in milliseconds

Finally, it was investigated whether patterns in people's eye movements could predict logicality in their judgements. A Pearson's correlation was computed to assess the relationship between refixation dwell times and logicality when judging conjunction probability. No relationship was found between statement refixation dwell time and logicality, r = -0.11, n = 41, p = .498; however, there was a moderately negative relationship found between description refixation dwell time and logicality, r = -0.36, n = 41, p = .020. A scatterplot summarises the results (see Figure 3. 14). Overall, the amount of time spent refixating on the statements made no difference to the logicality of people's conjunction judgements; however, it was found that the longer people refixated on the descriptions, the less logical their judgements were. This supports the deliberative heuristic idea. Perhaps when people reflect on their answers (e.g., search for the answer in the description) that is when the heuristic biases their judgements.

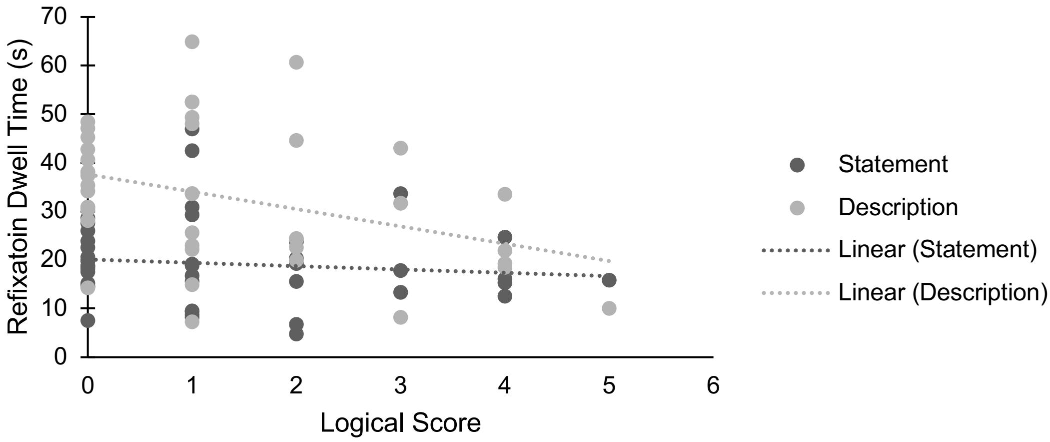


Figure 3. 14. Relationship between logicality on the conflict statements, and refixation dwell times (in seconds) for statement and description AoIs

Discussion

Support of Hypotheses and Interpretation of Results. The aim of this experiment was to uncover whether individual differences in patterns of eye movements could account for differences in individuals' propensity to generate logical probability judgements and focussed in particular on the question of conflict detection between heuristic and logical considerations. A secondary objective of this research was to determine how individual differences impact on the logicality of individuals' probability judgements. Experiment 1 tested the hypothesis that individuals who tend to engage more naturally in effortful thinking are also more prone to produce illogical probability judgements if they have time to deliberate because their deliberation is biased by non-logical heuristics. Specifically, Experiment 1 examined the impact of time pressure on the sensitivity to the conflict between heuristic and logical assessment in social judgements under uncertainty. The results supported the hypothesis and revealed that in addition to people being able to detect conflict between heuristic and logical considerations, it appeared that they were also able to have logical intuitions. In other words, their intuitive responses were also portraying an understanding of logic; they were not simply applying the representativeness heuristic blindly.

Experiment 2 was designed to replicate Experiment 1's findings, but also to move away from time pressure as a manipulation and instead use a more fine-grained methodology. It employed the use of an eye-tracker to test the hypothesis that in situations of conflict between representativeness and logicality (R/I and U/L statements), individuals who provide logical answers will show different patterns of eye-movement disruption than individuals who provide heuristic answers. The aim of this experiment was to shed light on the processes involved in generating conjunction probability judgements. Specifically, it focussed on whether eye movements would differ as a function of judgment setting: conflict detection between heuristic and logical considerations; and provided empirical data to test the widely accepted assumption that heuristic thinking is fast and automatic while logical thinking is slower and more deliberate.

In light of Experiment 1's findings, it was unclear how the behavioural data in Experiment 2 would unfold as differences were found in the reaction time data, but not the heuristic scores data. In the first experiment, in the absence of time pressure, there was no evidence of conflict detection in the heuristic responses between no conflict and conflict statements; however, there was a difference found in the reaction times. Participants took longer to answer conflict statements logically compared to heuristically or when there was no conflict present. There was also no difference in reaction times between conflict-heuristic judgements and no-conflict judgements. However, the heuristic scores data presented no evidence of conflict detection in the unlimited time condition in Experiment 1. Thus, the question remained whether this lack of difference in responses in fact meant that people were unable to perceive the conflict. Hence the collection of eye-tracking data to test this. In accordance with Experiment 1, there was no evidence found of conflict detection in the behavioural data in Experiment 2. The pattern of results is in fact very similar to the unlimited time condition in Experiment 1 (E1: $M_{\_noconflict} = 90.25$, $M_{\_conflict} = 87.00$ vs. E2: $M_{\_noconflict} = 86.59$, $M_{\_conflict} = 83.23$), which is unsurprising as both tested participants under similar conditions (i.e., unrestricted time to complete the task). In the absence of time pressure behavioural data indicates no sensitivity to conflict, but on examination of the eye movement data, this was in fact not the case. The eye-movement data revealed evidence for conflict sensitivity. Participants redirected their eyes more often towards the description in the absence of conflict; however, they redirected their eyes more often to the statement in the presence of conflict. Sympathetic to these findings, participants refixation dwell times were longer when looking at the description when conflict was absent, but when conflict was present their refixation dwell times were longer on the statements. These findings provide strong evidence for conflict detection as eye movement are difficult to censor due to their speed and automaticity (Glaholt & Reingold, 2011). Thus, they give reliable insight as to the cognitive processes of judgements and decisions. One caveat in using eye tracking as a process tracing measure is that it is assumes that the decision maker's attention is focused at the point of fixation. But there is evidence that people are able to redirect their attention undetected to different areas of the visual field away from their point of gaze (Posner, Snyder, & Davison, 1980). However, according to Glaholt and Reingold (2011) during natural viewing, attention and eye movements are tightly coupled. The "focus of attention tends to shift to a new location just prior to a shift in gaze to that location and consequently the spatial distribution of eye fixations is a good indirect measure of the distribution of visual attention" (Glaholt & Reingold, 2011, p. 127). If it is true that fixation gaze is a good measure of attention, and that automatic, superficial levels of processing have shorter fixation latencies, while deeper processing and a deliberate consideration of information have longer fixations (Glöckner & Herbold, 2011; Velichkovsky et al., 2002), one can speculate that the results from Experiment 2 show that when conflict was present in the trial the participants spent longer fixating on (i.e., their attention was directed towards) the statement; however, when conflict was absent this attention was focused on the description. These results show that people use different information search strategies depending on whether conflict is present or absent in the trials. This evidence of conflict detection provides support for the parallel processing model of the two processes. The two systems are activated from the beginning and work in parallel with one another, allowing a person to easily detect conflict when they each produce a different response. The fact that conflict detection can be witnessed in our eye movements (e.g., uncensored, extremely fast and automatic) suggests that people are extremely well adapted to detect conflict. This goes against the default interventionist account. A serial model of reasoning that states people are very bad at detecting conflict, and when they do they lack the motivation and resources to inhibit the intuitive response.

The evidence presented for the logical intuitions and heuristic reflections accounts are unfortunately inconclusive in this study as judgement latencies yielded a not significant correlation (r = -0.28, p = .079). A major reason for this could be due to the study being underpowered. The study was originally powered for an eye tracking study and 41 participants was enough to show differences in eye movements; however, it was unfortunately not high enough to show differences in the behavioural data. A posthoc analysis revealed that the sample of 41 participants only generated $1 - \beta = 0.28$ power. To reach a medium effect size (d = 0.5) and $1 - \beta = 0.80$, I should have tested 68 participants. Although I found evidence for conflict detection in the eye movement data; there was no evidence of logical intuitions in the behavioural data. However, there was a pattern emerging in the correlational results and perhaps with a larger sample this would have reached statistical significance. A negative relationship between logicality and total duration of the task was found. In other words, the longer participants took to complete the task, the less logical, or the more in line with heuristic considerations their answers were. This suggests that the faster judgments were more logical, which supports the assumption that logical answers come first and are intuitive but are overridden by the compelling heuristic answer when people take time to deliberate. When this happens, logical responding requires even more deliberation and effort to be reinstated. However, these findings were not significant, they were merely an emerging pattern in the data. Further testing should be done using an appropriately powered sample in future to consolidate these conjectures.

Generalizability of Results. If people are able to detect conflict in conjunction probability judgments and they also have logical intuitions, the question remains as to why then do the majority of people still commit conjunction fallacies? A possible explanation for this could be that although people are able to generate fast logical intuitions about conjunctive probability judgments, when they detect conflict they employ deliberate thinking, and this is when the logical intuitions are overcome by slow heuristic deliberations. De Neys (2012) and Handley and Trippas (2015) offer explanations as to why this might be the case. When people fail to make normatively correct judgements, there are often implicit measures such as eye movements or response times that reveal a sensitivity to the normative information, even though this does not translate into correct answers. Additionally, when people make normatively correct judgements they often do not mention the critical aspects of the normative solution in their verbal protocols. These findings have led to the idea that logical judgements can be activated implicitly and thus perhaps outside of awareness, which is characteristic of intuitive rather than deliberative reasoning (Trippas, Handley, Verde, & Morsanyi, 2016). Nonetheless, despite the convincing evidence for conflict detection presented in the eye movement data in this study, it remains true that the heuristic answer stayed prevalent even under conflict. Another possible explanation for the high heuristic responses in this study could be a methodological artefact. The Eyelink II eye tracker has a strict requirement for head-stabilization during recording (e.g., resting the chin on a chin-rest while the forehead is held up against a forehead-bar). The participants' heads were strapped into an intimidating looking eye-tracking machine while their chins rested on a chin bar to reduce head movement, which could have induced a more thoughtful deliberative mindset. If indeed logical answers come first and are intuitive, but are overridden by the compelling heuristic answer when people take time to deliberate; then this restrictive posture might be causing people to deliberate and overthink their answers. No instructions were given to the participants as to how much time they should take to make their judgements; however, being strapped into such serious scientific equipment might have also unknowingly prompted them to think harder about their responses and in doing so increased their deliberative heuristic responses. One way to test this moving forward, might be to use a less intrusive eye-tracker for any future studies. Current video-based eye tracking technology allows for head-free eye movement monitoring, which is "idea for observing computer-based decision" (Glaholt & Reingold, 2011, p. 128). For example, screen-based eye-trackers are unobtrusive yet robust measurers of gaze and fixation data. The cameras are located on the computer screen; thus, they are far less intrusive than the Eyelink II eye tracker. There is no heavy camera contraption strapped to the participants heads and no stabilization bar to rest the chin on.

Another piece of evidence that points to a disconnect between logicality and deliberative thinking in conjunction fallacies is the finding that working memory capacity did not correlate with logicality. Stanovich and West (2008) conducted seven different studies and observed that a large number of thinking biases were uncorrelated with cognitive ability. They state that "people of higher cognitive ability are no more likely to recognize the need for a normative principle than are individuals of lower cognitive ability" (p. 690). However, when told what the bias is and when they needed to avoid it, people with higher cognitive ability displayed fewer reasoning biases than those with lower cognitive ability. The authors hypothesize that early in one's developmental history the relevant mindware is not present, thus conflict detection is absent and the heuristic response is inevitable. As experience with statistics and probabilities grows, so does conflict detection because the relevant mindware is now available to engage the analytic system and override the heuristic response. Finally, they propose "that the mindware used in analytic simulation becomes so tightly compiled that it is triggered in the manner of a natural heuristic response" (p. 690).

The evidence presented in this study goes against the default-interventionist interpretation since participants' eye movements suggest they easily detected the conflict between logical and heuristic considerations. When conflict was present in the CCPJ task, participants refixated their eye movements more often and also had longer refixation dwell times on the statement AoIs; however, in the absence of conflict they refixated more often and for longer time periods on the descriptions. This result suggests that people are able to detect when conflict is present in the CCPJ task statements, and they are employing different strategies to solve the probability judgements dependent on whether conflict is present or absent in the trial. When conflict is present they are concentrating more of their attention searching for answers in the statement, whereas when conflict is absent they search for answers in the descriptions. This evidence of conflict sensitivity better suits a parallel-competitive account of reasoning. This account believes that both systems of reasoning operate in tandem until conflict arises. Then it is possible for System 2 to override System 1; however, this does not always happen due to the heuristic response being so persuasive. The results showed a drop in heuristic responses for the conflict statements compared with the no conflict statements, but this difference was not significant. This suggests that the heuristic responses might have been too convincing to override. Or perhaps the participants lacked the motivation to inhibit them due to task fatigue experienced by wearing the head mounted cameras as discussed previously.

Final Remarks and Future Research. The eye movement data showed that individuals are able to detect the conflict between logical and intuitive considerations in conjunction judgements. However, the behavioural data in this study gave no indication as to the mode of thought used when making the judgements. A future consideration might involve combining eye tracking with a verbal protocol procedure to better understand the thought processes taking place while making judgements. Albertazzi (2006) suggests combining the use of eye fixations and concurrent verbal protocols to illuminate the processes behind observing and describing an object because drawing conclusions using both methods is far more succinct than only using either eye movements, or verbal protocols alone. Verbal protocol involves participants describing what they are doing or thinking (i.e., “think aloud”) either while they are making a judgement, or retrospectively after already having made the judgement. While this is very useful if you want to know the strategies used to make the judgement, there are also shortcomings. Concurrent verbalisation (e.g., simultaneous verbalisation while making a judgement) is effectively a secondary task that places burden on the cognitive resources and has been shown to affect decision accuracy (Russo, Johnson, & Stephens, 1989). While retrospective verbalisation relies on the participants’ having an accurate memory of how the decision or judgement unfolded and has been shown to reflect substantial forgetting and confabulation (Russo et al., 1989). Although, watching one’s eye movements while performing a retrospective verbalisation might work well, as the video of the eye movements may help to jog the participants memory.

The average time taken to complete the CCPJ task was 219.65 seconds (*SD* = 61.10), which equates to 13.73 seconds per trial. Most adults read at a rate of approximately 300 words per minute and this can increase up to 450 words per minute for college and university students. The average number of words across the 16 vignette descriptions was 37.63 words, which means that between 5 and 7.5 seconds were spent reading the description, leaving 6.23 to 8.73 seconds to make a judgement. These are quite long response times and more than enough time to inhibit a heuristic response. However, the average number of logical (i.e., correct) judgements made was less than 2 (M = 1.32, SD = 1.42) out of a possible eight. The participants must be assumed to have sufficient cognitive ability to perform academically as they all gained entrance to university. These results suggest that even with the cognitive capability and enough time to employ deliberation, the majority of participants are still inhibiting the logical response (e.g., assuming that intuitive and deliberate processes run in parallel) on conflict statements. In line with the original hypotheses of this experiment, and the assumption of logical intuitions and heuristic reflections, a possible explanation could be that logical answers come first and are intuitive but are overridden by the compelling heuristic answer when people take time to deliberate (which they did in this study, taking almost 9 seconds in some cases to make an answer). When this happens, logical responding requires even more deliberation and effort to be reinstated, which would explain the low number of logical judgements made for the conflict trials. If deliberation, or in this case heuristic thinking, is actually resource intensive then one might expect that loading cognitive resources will actually allow for more logical answers. This is exactly what was found in Experiment 1: the impact of time pressure increased logicality. It might sound counterintuitive, however overloading cognitive resources will also deplete one's ability to deliberate, thus inducing intuitive thinking and increasing logic. However, this assumption needs further testing and it is important to note that the reason for most people making errors under conflict might simply be that people are uncertain about their answers under conflict, so they guess because they are forced to do so by the task. To further test the validity of the idea that heuristic answers are a result of deliberative processing and logical answers are derived from intuitive processing, I will next consider employing manipulations that decrease cognitive functioning as a means to increase intuitive processing. Evans, Dillon and Rand

(2015) warn researchers against the use of response latencies to infer reasoning type/process, hence the substitution of time pressure for alternative methods of encouraging intuition. To be more specific, I will test whether employing a task that either depletes or overloads working memory has any effect on people's logicality when solving conjunction judgements. By removing access to cognitive functions such as working memory (System 2), one would supposedly increase reliance on System 1 processing.

Chapter 4: Inhibiting Deliberative Thinking to Increase the Logicality of Conjunction Probability Judgements

The traditional dual-process theories of reasoning assume that humans have two distinct reasoning systems (e.g., Stanovich & West, 2000; Evans, 2003). System 1, also known as the intuitive system, is believed to be the default human reasoning system. It operates quickly and automatically by relying on prior knowledge and beliefs. System 2, also known as the analytic system, is slower, more deliberate, uses logical principles and is extremely taxing on executive resources (i.e., computational resources such as working memory). Assuming that System 2 reasoning is heavily demanding and draining on cognitive resources, then by depriving participants of these resources we potentially eradicate System 2 during the initial response phase. Methods to do so, might include overloading cognitive resources with a dot memory task (e.g., Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; De Neys & Schaeken, 2007), where individuals are required to keep a piece of information stored in their working memory while they simultaneously solve a conjunction judgement. This action of solving dual tasks places a burden on cognitive resources allowing for less available cognitive capacity for the secondary task. With less available resources, one assumes that the individuals will rely more heavily on intuitive processing to solve the conjunction judgement. Another option to test the model of "heuristic answers as deliberative processing" could involve depleting ego (i.e., self-control and cognitive energy) with an *e*-task (Sripada, Kessler, & Jonides, 2014). The limited-strength theory assumes that people have a finite capacity for performing tasks of self-control, called ego. Once this storage of energy is depleted, it is assumed to adversely affect subsequent tasks of self-control because cognitive resources have been drained or depleted (Myers, et al., 2018). Thus, depleting individuals' energy, or ego, before completing a conjunction judgement task would also induce the use of intuitive processing on that task as there are less available resources for deliberation.

Experiment 3: The Influence of Ego Depletion and Impulsivity on the Conjunction Fallacy

People are able to exhibit self-control in their lives as witnessed through their behaviours. Some obvious ones include refraining from urges such as overeating, drinking too much alcohol, or acting violently against someone when provoked (Tice & Bratslavsky, 2000; Vohs & Heatherton, 2000; Muraven, Collins, & Neinhaus, 2002). However, not all human behaviour involves deliberate self-control. It has been shown that some behaviour is influenced by automatic processes. Goal oriented behaviour can indeed be activated out of awareness and function nonconsciously to guide self-regulation effectively (Bargh, Chen, & Burrows, 1996; Bargh, et al., 2001). Nevertheless, the portion of behaviour that requires volition is "important to the long-term health, happiness and success of the individual", plus people have the ability to deliberate, think consciously about an issue and offer controlled responses of the self (Baumeister, Bratslavsky, Muraven & Tice, 1998, p. 1252). However, despite this remarkable capacity to regulate the self and overcome tempting desires, many behavioural and social problems arise from persistent lapses of self-control (Hagger, et al., 2010). This suggests that people's ability to exert self-control over their conscious behaviours is limited, and that without a store of energy for volition, people are not able to implement deliberate goal-oriented behaviour. The idea that volition expends a limited source of energy (i.e., ego) is not a new one. Indeed, Freud stated that "[Thought-processes] represent displacements of mental energy which are effected somewhere in the interior of the apparatus as this energy proceeds on its way towards action" (1923/1961, p. 19). He believed that the ego required some form of energy to accomplish its tasks and resist the energy depleting promptings of the id and superego (Baumeister, et al., 1998).

Ego depletion. The *limited-strength model* of self-regulation states that "people's ability to self-regulate is governed by a limited pool of resources; such that repeated acts of self-regulation deplete this resource pool leaving the participant vulnerable to subsequent self-regulatory failure until this resource pool is replenished" (Myers, et al., 2018, p. 2). Self-regulation, or self-control, entails "overriding or altering a predominant response tendency" (Schmeichel & Vohs, 2009, p. 770). This proposed depletion of self-regulatory energy with repeated acts of self-regulation is known as the *ego-depletion effect*. It can be likened to a muscle that tires after consistent use and is rendered useless until given time to rest and replenish (Baumeister, et al., 1998; Muraven, Tice & Baumeister, 1998). In 1998, two seminal articles emerged; the first by Baumeister and colleagues and the second by Muraven and colleagues, which provided empirical evidence supporting the limited-strength model. In addition to this, they also established the sequential-task paradigm as the preferred method for exploring the limited-strength model. This paradigm refers to the series of experimental manipulations that are presented to the participants. This usually entails two sequential tasks: an intervention task and an outcome task. The dependent variable is assessed by performance on the outcome task, which always involves some form of self-regulation, and is the same across both the experimental and control conditions. The intervention task varies across conditions; the control condition does not require any self-regulation but the experimental condition does. This task can vary vastly but must rely on a domain-general pool of resources. For example, having to resist the urge to eat chocolate and subsequently having to solve unsolvable puzzles (Baumeister, et al., 1998). The authors found that those who had to expend self-regulatory energy (e.g., resisting eating the chocolate) also gave up solving the puzzles faster than those who had not resisted the chocolate. This led the authors to conclude that an action such as self-regulated eating exhausted the general resource pool needed for persistence on the outcome task. A wide variety of tasks involving self-control have been utilised in the literature to induce ego-depletion. Ranging from eating inhibition, to inhibiting feelings while watching a disturbing clip, to letter crossing tasks. All of these result in a poor performance on a subsequent unrelated self-control task (Baumeister et al., 1998). This is due to the fact that the depleted participant tends to rest in the subsequent task and are thus more likely to act on impulse (Baumeister et al., 1998). Feeney, Shafto & Dunning (2007) state that individuals who are depleted of System 2 resources are unlikely to suppress the incorrect response from System 1 in favour of System 2, thus giving an intuitive and incorrect response.

In their seminal paper, Baumeister and colleagues (1998) questioned the toll that volition takes on the self. They suspected that acts such as judgement and decision making, responsibility taking, initiating and inhibiting behaviour, and making and executing plans, depleted a common inner resource called "ego", and they conducted a series of experiments that tested this. They hypothesised that acts of volition and self-control would cause ego depletion. More specifically, following an initial act of volition, there would be less of this energy resource available for subsequent acts. They tested this hypothesis across a series of experiments examining a variety of seemingly unrelated acts: impulse control, attitudinal choices, affect regulation, and response regulation. The first experiment required hungry participants to resist the temptation of chocolate and perform a much less desirable action of eating radishes (i.e., impulse control) followed by persistence at a frustrating unsolvable puzzle task. It was found that participants who forced themselves to eat radishes rather than chocolate subsequently quit faster on the puzzle tasks than participants who had not had to exert self-control over eating. Thus, resisting temptation had produced a "psychic cost", which resulted in less ability to persist when faced with frustration (Baumeister et al., 1998, p. 1255). The second experiment explored whether the same resource depletion was present in acts of choice. Participants were informed they had to read a speech. They were either told which speech they had to deliver (no choice) or they were given a choice of speeches (high choice). Speeches were either pro-tuition fee raise (counterattitudinal), or anti-tuition fee raise (proattitudinal). These choices were followed by persistence at the same unsolvable puzzle task as in the previous experiment. It was found that making a meaningful personal choice to perform attitude-relevant behaviour caused a similar decrease in persistence as found in the first experiment. Hence, taking responsibility for a counterattitudinal behaviour depleted ego resources, which revealed that different types of volitional actions have the same depleting function on ego. These two experiments both showed that volition expended ego, which resulted in less persistence on unsolvable problems (Baumeister et al., 1998). Some believe that persisting in the face of frustration requires an act of self-control (overcoming the impulse to quit); however, others believe that it is adaptive to quit early on unsolvable problems. Persistence is only productive if it leads to success, thus squandering time and effort on an unsolvable problem is wasteful. Following on from this idea, the authors designed a third experiment using affect regulation (i.e., controlling one's emotions) and solvable problems. This study instructed participants to either suppress or not to suppress their emotions, while watching either a funny or sad 10-minute video clip. Following this they completed an anagram task. Anagrams were chosen because they are solvable puzzles, which require people to keep breaking and reforming combinations of letters and persist through many initial failures; a task that requires self-control. Results showed that supressing emotions (both happy and sad) led to a decrease in performance of solvable anagrams, which indicates that acts of volition do in fact decrease persistence and performance on both solvable and unsolvable puzzles. Thus, the idea that people quit early on unsolvable puzzles because it is adaptive to do so cannot be accounted for. Instead these results support the idea that a valuable and limited resource of self is depleted by an initial act of volition. The fourth and final experiment was designed to complement experiment 2. Instead of showing how an act of responsible decision making undermined subsequent self-regulation, this experiment showed how an act of self-regulation undermined subsequent decision making. Participants initially completed a regulatory-depletion task (the letter-*e* task) in one of two conditions: control or depletion. The control condition required participants to read a page of meaningless text and simply cross off all instances of the letter *e*. The depletion condition was more difficult and required them to conform to rules and withhold a prepotent response to the stimulus. They were told to cross off the letter *e*, but should withhold this response if the *e* was next to or one extra letter away from another vowel. For example, they would not cross off the *e* in the word *vowel* because it is one letter away from another vowel, *o*. After completing the letter-*e* task, participants were shown a boring movie clip. They were cautioned to watch as long as necessary to understand the content of the movie and answer a few questions afterwards, but they did not have to finish the movie and it was up to the participant when to stop. Half of the participants terminated the movie by pressing down on a button (active quit condition) while the other half lifted their hand off a button (passive quit condition). Results showed that high self-regulation (ego-depletion) caused people to be more passive (i.e., more prone to favour the passive-response option). Participants who had to do less work to quit (i.e., lift hand off a button) were more inclined to quit sooner when they were depleted than when they were not. Participants who had to initiate an action to quit (i.e., press a button down) tended to watch the movie longer when they were depleted than when they were not. In the absence of ego depletion there was no difference in time spent watching the movie. These findings provide further support for ego-depletion, such that ego depletion increased subsequent passivity. In other words, depleted people were less inclined to make active responses and were more inclined to choose the easier route, "as if carried along by inertia" (Baumeister et al., 1998, p. 1261). Overall, this investigation suggested that people do have a limited resource of energy, or ego, available to them, and that this resource is expended or depleted through the use of a wide range of different actions, such as self-control, responsible decision making, and active choice. All four experiments provided support for the notion that volition causes ego-depletion. Specifically, following one initial act of volition there is less resource available for subsequent acts.

Muraven, Tice and Baumeister (1998) also presented a series of experiments as empirical evidence for the limited-strength model of self-regulation. They conducted three experiments in which participants engaged in self-regulation followed by a subsequent seemingly unrelated task also requiring self-regulation. The first experiment found that emotional regulation (i.e., overcoming a current emotional state and replacing it with a different one) to an upsetting movie resulted in a decrease of physical stamina, as measured by how long participants squeezed a handgrip. Experiment 2 replicated Experiment 1 but used very different methods. The authors found that self-regulation by suppressing forbidden thoughts (i.e., told to NOT think about a white bear as much as possible) resulted in a subsequent tendency to give up more quickly on unsolvable anagrams. Both these experiments provided support for the notion that self-regulation involves a limited capacity of resources that can become temporarily depleted. Experiment 3 again sought to replicate 1 and 2, but also rule out several alternative explanations. The authors found that supressing thoughts (using the same thought suppression manipulation as in Experiment 2) impaired subsequent efforts to control the expression of amusement and enjoyment in response to a humorous video. The authors again replicated the finding that self-regulatory performance drops off after a prior attempt at self-regulation. Participants who regulated their thoughts by not thinking about a white bear, were also less successful at controlling their emotions subsequently. Finally, a fourth study was conducted using autobiographical accounts that linked prior self-regulatory demands and indicators of regulatory capacity with poorer (or better, depending on the indicator) self-regulation of emotion.

Gailliot and colleagues (2007) tested a biologically based intervention to restore the self-control resource. They discovered that acts of self-control not only depleted ego, but also depleted glucose levels (one of the body's main sources of energy). And they found that low glucose levels correlated with poor performance on self-control tests. Remarkably, by getting their participants to drink sugary lemonade, they restored glucose levels to the bloodstream, which also restored the capacity for self-control. Interestingly, research conducted by Schmeichel, Vohs, and Baumeister (2003) suggested that ego-depletion led to poorer intellectual performance, but only on higher order, complex tasks such as logical reasoning. They propose that the more complex forms of reasoning require active self-control and executive functioning, which consume ego (or energy) reserves, because these tasks were impaired following ego-depletion. While simpler information processing, such as general knowledge or a vocabulary test, were unaffected by ego-depletion.

Hagger et al. (2010) provided strong empirical support for the limited-strength model through a meta-analysis of 83 studies. They authors concluded that "the strength model is a useful explanatory system with which to understand self-control, but further refinements may be necessary, particularly when it comes to the identification of mechanisms" (Hagger et al., 2010, p. 520). This work proposed that the ego-depletion effect was reliable and represented a medium to large effect size. However, in 2014 Carter and McCullough reanalysed the data, but this time corrected for small-study effects and found strong signs of publication bias and an indication that the depletion effect is not meaningfully different from zero. The publication bias causes the sample to become unrepresentative and is prevalent in meta-analysis as the majority of studies chosen for the analysis are published studies, thus publication status is influenced by the study's results. In other words, studies with significant and theory-supporting results tend to be more likely to end up in meta-analytic samples (Carter & McCullough, 2014). In addition to questions raised by Carter and McCullough (2014), recently there has been some conflicting evidence for ego-depletion including work that has questioned the strength of the ego-depletion effect and challenged whether it exists at all (Hagger et al., 2016). Some research suggests that applying self-control does not always reduce subsequent self-control (Moller, Deci, & Ryan, 2006; Tice, Baumeister, Shmueli, & Muraven, 2007; Job, Dweck & Walton, 2010). For example, in a study conducted by Muraven and Slessareva (2003), it was shown that participants who were motivated by incentives to apply self-control, did not exhibit any signs of ego-depletion on subsequent tasks. Across a series of three experiments, the researchers determined that the effect of ego-depletion is moderated by motivation (i.e., incentives). In other words, a monetary incentive could compensate for lack of self-control resources. Indeed, when sufficiently motivated, ego-depleted participants performed as well on subsequent tasks as non-depleted participants. In a study conducted by Martijn, Tenbült, Merckelbach, Dreezens, & de Vries (2002) participants were told that performing a prior difficult task (i.e. controlling their emotions) would improve their performance on a subsequent task. They found that people's expectancies about diminished self-control moderated ego-depletion. Participants who had been informed that their performance would improve following the ego-depleting task, actually showed no decrease in performance on a subsequent self-control task of squeezing a handgrip. Schmeichel and Vohs (2009) conducted a series of experiments that showed sound evidence that self-affirmation counteracted ego depletion by promoting high levels of mental construal, which contributed to successful self-control. Job, Dweck & Walton (2010) addressed the question whether people's beliefs or implicit theories about self-control and willpower moderated the effects of ego-depletion. Their findings were similar to that above of Martijn et al. (2002), people who believed or were led to believe that they have a nonlimited resource of energy/ego, performed equally as well, or sometimes even better, on a subsequent demanding task. Indeed, a longitudinal study showed that the more people believed their energy/ego was limited, the poorer their self-regulation in the real world was when self-control demands were high. Finally, Xu et al. (2014) and Lurquin et al. (2016) both found non-significant results while testing for an ego-depletion effect, despite using reliable research methodologies and large sample sizes. These findings add to a growing body of evidence suggesting that ego-depletion is not a reliable phenomenon.

Ego depletion tasks. There are many different types of depletion tasks – both emotional and cognitive, but the problem is that most of these interventions are not evaluated for performance. Performance on the intervention task needs to be independently evaluated in order to show "a decrement in performance that would be indicative of a resource depletion"

(Myers et al., 2018, p. 3). This lack is often due to the fact that many of these tasks are not amenable to measurement (e.g., not thinking of a white bear). However, without demonstrating that intervention tasks do in fact deplete resources, the theory underlying the strength model remains an assumption. One of the most popular and commonly used depletion intervention tasks, which can be measured, is the letter-crossing task (see Hagger et al., 2010; Carter et al., 2015). It is a search and identification task usually involving the letter *e.* This task requires participants to locate the letter e according to different sets of rules. The first part of the task is designed to instil a habit and requires participants to cross out every letter *e* on a page of text. The second part of the task is the experimental condition and is designed to force participants to inhibit this habituated response. They are given a second page of text and told to cross out all instances of the letter *e* except in specific cases. For example, not if the letter *e* is next to, or one letter away from a vowel. This action of having to override a newly acquired habitual response is said to be the cause of resource-depletion (Tice et al., 2007; Xu et al., 2014).

The use of the letter *e* task requires that both the experimental and control conditions emit behavioural responses that can be tracked over time. "This is critical since the depletion of regulatory resources is a within subjects effect (occurs within participants over time) and the standard approach to observing it has been between-subjects using comparisons of experimental conditions" (Myers et al., 2018, p. 3). One of the only studies that has in fact measured changes across time is Arber et al. (2017) who presented participants with five different pages of text and asked them to mark off the letter *e*, except after certain rules. They found that across the five studies, participants' ability to detect target vowel combinations (e.g., *ie* and *ei*) declined as a function of time on task. These same participants also showed a decline in performance on a secondary task – working memory span. Finding a time dependent decline on the experimental task followed by a similar decline on a subsequent task is wholly in keeping with the strength model predictions. However, there are some limitations to the study. Rather than having an active control condition (e.g., crossing off instances of the letter *e*), the authors used a passive 10-min "chat" with the experimenter before completing the working memory span task. This study also did not have the habit forming first page. All five pages required the application of a rule. Indeed, some authors have argued that the habit-forming component of the letter e task is unnecessary, and that the immediate implantation of both rules is enough to cause ego-depletion (Arber et al., 2017; Baumeister et al., 1998; Sripada et al., 2014; Hagger et al., 2016). These researchers argued that identifying the letter *e* in one set of cases and not in another set of cases is a self-regulatory action that in itself would require the use of limited resources; and this would be sufficient to cause resource depletion. The presence or absence of a habit-forming component in the letter *e* task has been up for debate recently. Baumeister and Vohs (2016) acknowledge that the Registered Replication Report (RRR) on ego depletion conducted by Hagger et al. (2016) had a methodological flaw, which was that the study did not start with a habit-forming stage prior to the instigation of a new set of rules. They see this as a reason to dismiss the null findings. "Without first instilling the habit, there is nothing to override. This may be a difficult cognitive judgment task, but no impulse is overridden, contrary to the nature of self-control tasks" (Baumeister and Vohs, 2016, p. 574).

Myers et al. (2018) offer two alternative explanation for the negative performance gradient found by Arber et al (2017). First, they suggest that by following the directions to inhibit an *ie* or *ei* response participants must override an automatic response to read as these vowel pairs are embedded in text. If this is indeed the case, then the depleting effect would be equal across the experimental and control conditions, as both these conditions require the overriding of the automatic reading response. Secondly, the authors suggest that simply following instruction to obey multiple rules could be the cause of depletion as by increasing the task difficulty, one also increases cognitive load. The fact that the subsequent task involves behavioural inhibition is incidental and might not contribute to the depletion effects.

If this is true, and the letter-*e* task manipulations are simply increasing the task complexity and difficulty, then the declining performance on the second task is warranted: "obviously, a more difficult task leads to poorer performance" (Myers et.al.,2018, p.4). However, task difficulty cannot account for the declining performance over time. A more difficult task would mean poorer results initially, but over time one would not expect to see greater deficits in performance. In fact, one might expect to see performance either remain constant or improve due to practice effects as one does in difficult skill-based tasks such as Stroop tasks. Practice effects are not observed in the letter *e* task because it is a fairly easy task in general. It might take some effort to complete, but it is not difficult and does not require much skill to complete. Myers et al. (2018) believe it is these characteristics that make the letter e task ideal for investigating ego depletion.

Barratt Impulsiveness Scale. Impulsivity has been found to be a strong predictor of certain behaviours, for example gambling (Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007), demonstrating intelligent professional conduct (Costa & Kallick, 2000), and exhibiting low conscientiousness but high neuroticism (Bamberger & Bacharach, 2006). Dickman (1990) defined impulsivity as the tendency to deliberate less than most people of equal ability before acting. This link to less deliberation prompted me to further explore impulsivity's relationship to thinking dispositions by using the Barratt Impulsiveness Scale version 11 (BIS-11; Patton, Stanford & Barratt, 1995). Impulsivity is a personality dimension described as "acting without thinking" (Moeller et al., 2001). While Martin and Potts (2009) believe impulsivity may be a multidimensional construct, where individuals lack the ability to inhibit certain behaviours and seek immediate rewards. In line with these definitions, one could assume that impulsive people tend to rely more on their intuition, while non-impulsive people prefer to deliberate their judgements. Patton and colleagues (1995) have proposed that impulsivity is comprised of three higher-order factors. The first, attentional impulsiveness, is the ability to focus on the tasks at hand and cognitive instability. An item example of attentional impulsiveness is “I get easily bored when solving thought problems”. The second factor is motor impulsiveness, which involves acting on the spur of the moment and perseverance (e.g., I do things without thinking). The third and final factor is non-planning, which includes self-control and cognitive complexity (e.g., I am more interested in the present than the future). The authors report internal consistency coefficients for the BIS-11 total score that range from 0.79 to 0.83 for separate populations of undergraduates, substance-abuse patients, general psychiatric patients, and prison inmates. The results of this study (Patton et al., 1995) conclude that the total score of the BIS-11 is an internally consistent measure of impulsiveness.

The current study. The objective of this study was to test the hypothesis that decreasing individuals’ capacity for effortful thinking will also have a positive impact on probability judgments because it will impede the inhibition of the rapid, logical response by conscious effortful deliberation. Using the Comparison Conjunction Probability Judgement (CCPJ) task methodology, this study aims to identify the impact of ego depletion of cognitive resources on the acceptance rates of both logical and illogical statements. It will provide the first assessment of the impact of cognitive depletion on the relative weight of heuristic and logical considerations in probability judgments. Additionally, this study will explore whether individual differences such as the propensity to be impulsive has any bearing on one’s logicality when judging conjunction probability judgements. In accordance with the assumption that impulsivity reflects fast, intuitive, System 1 thinking, I expect to find that the higher individuals score on the BIS-11 (i.e., the less impulsive they are), the more heuristic they will be in their judgements on the CCPJ task. Furthermore, and in line with the counterintuitive assumption that heuristic thinking in fact leads to more logical answers, I expect to find that individuals who assess themselves as being impulsive, also exhibit higher logicality in their conjunction judgements, as they are less prone to deliberating and being biased by the deliberative heuristic.

Method

Participants. Guidelines suggest that for analysis of variance (ANOVA) designs in which factors have medium to large effects, a sample size in the range of 30 participants per cell yields approximately 80% power (Van Voorhis & Morgan, 2007). Ninety-eight psychology students from Kingston University were recruited for this experiment through the Sona-System – a shared participant scheme. Two students were already familiar with the task and another was completely unbiased in his/her answers, thus these three participants were removed from the sample. Of the remaining 95 participants (18 men and 77 women; mean age = 22 years, *SD* = 5.49), 23% were postgraduates and 77% were undergraduates. The majority of the participants were enrolled in psychology courses (85%); while the remainder were studying subjects including business (1%), chemistry (1%), creative writing (4%), dance (1%), film (2%), Law (2%), marketing (1%), property (1%), sustainability (1%) and theatre (1%). The experiment was conducted in English; 82% of the participants were fluent, native English speakers, 14% were proficient, 3% were intermediate and 1% had basic knowledge of the language. Participants were tested in isolation in a laboratory. Participation was voluntary and anonymous. The study took approximately 30 minutes to complete, and all participants were paid £4 and received 30 SONA credits for their time and participation.

A control group of 50 participants was added to this sample from the unlimited condition in experiment 1. In this condition, participants simply completed the 16 CCPJ trials in their own time and capacity. They were not exposed to any time pressure nor were they put under any cognitive load. This brought the total sample size to 146 participants. In the control group, (8 men and 42 women; mean age = 24.84 years, *SD* = 9.47), 32% were postgraduates, 66% were undergraduates, and one participant did not disclose. The majority of the participants were studying psychology (86%) and the experiment was conducted in English;

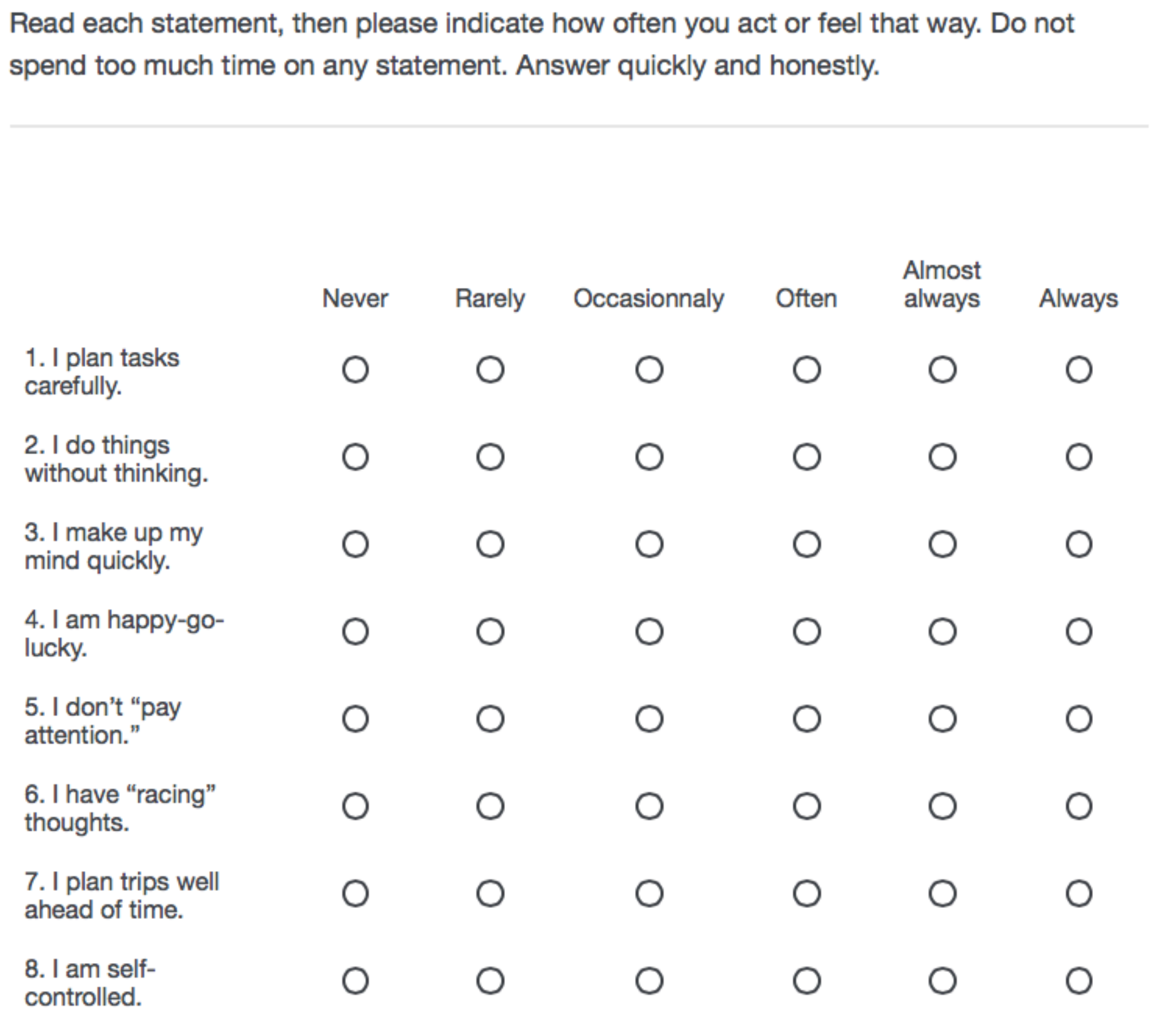
64% of the participants were fluent, native English speakers, while the remaining 36% were second or third language English speakers.

Materials. *The letter-e task* used in this experiment was the same version used by Sripada and colleagues (2014). The Registered Replication Reports (RRR) in conjunction with the Open Science Framework (OSF) and in an ongoing effort to publish collections of replications based on a shared and vetted protocol has made available to the public the same letter-*e* task used in the Sripada et al. (2014) study. The letter-*e* task was selected for the function of ego depletion because it has yielded some of the highest effect sizes in inducing regulatory depletion (Hagger et al., 2010). It is a modified version of the letter-*e* task (Baumeister et al., 1998) lasting 7.5 minutes and consisting of 150 trials. Participants first completed a practise version (20 trials) of the letter-*e* task before they completed the main version. There were two conditions: control and experimental. Words appeared on the screen one at a time, for 1500 milliseconds each. In the experimental version of the task participants were required to suppress a prepotent tendency. They were instructed to press the "1" key when a word with the letter *e* was shown, but they had to withhold this response if the *e* was next to or one extra letter away from another vowel. Vowels included: "a, e, i, o, u, and y". The control version was matched in all respects, except that participants pressed the "1" key whenever a word with the letter *e* was shown, regardless of where in the word the *e* appeared. Consequently, no suppression of prepotent tendencies was required. Both conditions of the letter-*e* task were administered using E-Prime 2 software, which recorded participants' reaction times and errors for the tasks.

The Comparative Conjunction Probability Judgement Task (CCPJ) was the same task administered to participants in the unlimited time pressure condition of experiment 1. It was programmed in E-Prime 2 software and presented 16 CCPJ trials to participants on a computer screen. Each CCPJ trial consisted of a thumbnail description and a statement sentence. The descriptions were short paragraphs that outlined the personality traits, hobbies and vocations of fictitious people intended to be congruent with a stereotype. For example, Bill was presented as an intelligent yet unimaginative, compulsive, and generally lifeless man, whom at school was strong in mathematics but weak in social studies and humanities. Each description was followed by a unique statement; one sentence that contained both a single clause and a conjunction clause. Statements always featured a category that corresponded to the stereotype suggested in the thumbnail description (e.g., Bill is an accountant) and an atypical category chosen to be at odds with the stereotype (e.g., Bill plays in a rock band for a hobby). The single and conjunction clauses were used to manipulate the conflict within the statements (see Table 2. 1, p. 74 for an example of conflict and non-conflict statements).

The Barratt Impulsiveness Scale – version 11 (BIS-11; Patton et al., 1995) is a questionnaire designed to measure impulsiveness. It is composed of 30 items describing common impulsive or non-impulsive (for reverse scored items) behaviours and preferences. For example: “I plan tasks carefully” and “I do things without thinking”. Items were answered on a 6-point Likert scale ranging from 1 = “never” to 6 = “always”. The BIS-11 was constructed in Qualtrics software and presented to participants electronically. See Appendix D for the complete BIS-11. This questionnaire was included to examine the relationship between impulsiveness and dual thought processes in conjunction judgements. A person who scores high in impulsiveness should in theory be more likely to use intuition when making conjunction judgements (Gerrard et al., 2008). If this is the case, then according to our hypothesis that intuitive considerations result in more logical conjunction judgements, those participants with higher BIS-11 scores should answer the CCPJ trials more logically than those with lower scores.

Design and procedure. The study used a 2 x 3 mixed design in which the dual-task paradigm was crossed with ego depletion manipulation using suppression of a prepotent tendency or no suppression. On arrival at the laboratory all participants were fully briefed on the study and gave their informed consent before starting the experiment. The first task to be administered was the BIS-11 questionnaire. This study was programmed in Qualtrics and completed on a computer screen. After reading the introduction, “People differ in the ways they act and think in different situations. This is a test to measure some of the ways in which you act and think”, the participants followed these instructions to complete the questionnaire, “Read each statement, then please indicate how often you act or feel that way. Do not spend too much time on any statement. Answer quickly and honestly”. The questionnaire items were presented in a matrix table and participants used a mouse to click their answers onto a 6-point Likert scale (0 = “never”, 5 = “always”, see Figure 4. 1 below for an example of the BIS-11 as displayed to the participants in Qualtrics software).



Read each statement, then please indicate how often you act or feel that way. Do not spend too much time on any statement. Answer quickly and honestly.

	Never	Rarely	Occasionnaly	Often	Almost always	Always
1. I plan tasks carefully.	○	○	○	○	○	○
2. I do things without thinking.	○	○	○	○	○	○
3. I make up my mind quickly.	○	○	○	○	○	○
4. I am happy-go-lucky.	○	○	○	○	○	○
5. I don't "pay attention."	○	○	○	○	○	○
6. I have "racing" thoughts.	○	○	○	○	○	○
7. I plan trips well ahead of time.	○	○	○	○	○	○
8. I am self-controlled.	○	○	○	○	○	○

Figure 4. 1. Screenshot of the Barratt Impulsiveness Scale (BIS-11; Patton et al., 1995) as displayed in Qualtrics Software

On completion of the BIS-11 questionnaire, participants were randomly allocated into one of the two letter-*e* task conditions - either low depletion or high depletion. In both conditions of the task, participants were required to engage in a brief practice period before the commencement of the experiment. The researcher first loaded the practice letter-*e* task onto the testing computer and read through the instructions with the participants ensuring they understood precisely what was required of them before starting the practice test (Figure 4. 2 demonstrates the instructions for both task conditions).

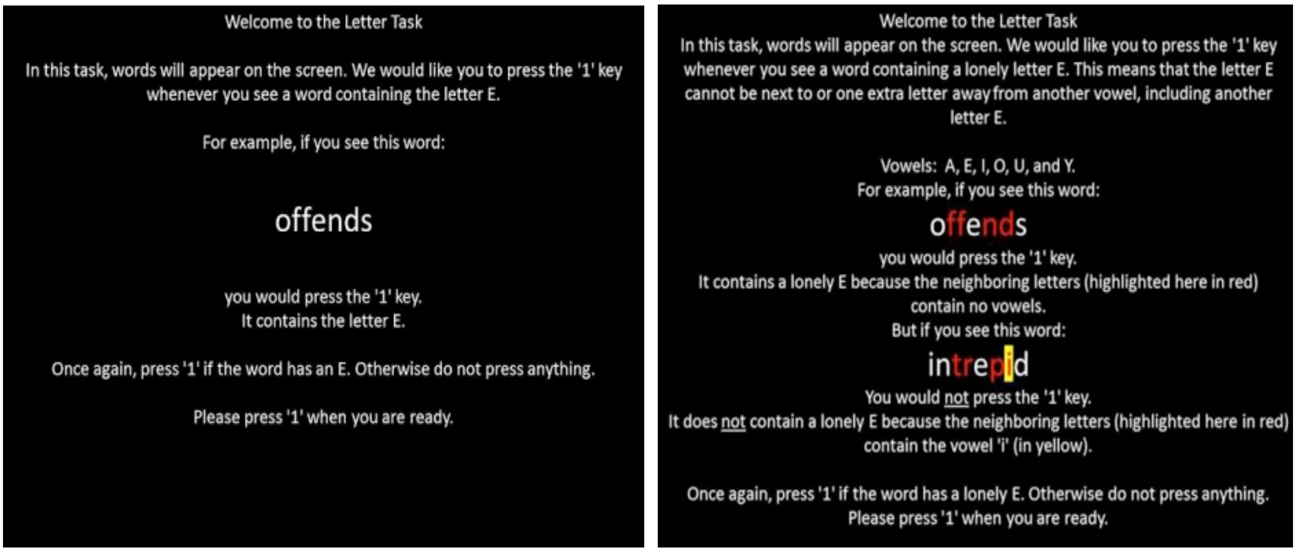


Figure 4. 2. Letter-e task instructions for the low depletion condition (left panel) and for the ego-depletion condition (right panel). *Note*, the control group was taken from E1 no time pressure condition

During the task, the experimenter left the room so there were no distractions for the participants. Participants were told to call the experimenter back into the room when they arrived at a screen saying, “The Task is Done”. Each trial consisted of a white word presented on a black background. Each word was displayed for 1500 milliseconds in Arial font sized 28. The words were preceded by a fixation cross, which was a white “+” sign on a black background, presented for 1500 milliseconds in Arial font sized 22.

On re-entering the room, the experimenter asked: “Did you fully understand the task?” If the participants did not understand, the experimenter repeated the instructions given previously for the relevant condition and answered any further questions. Once the participants confirmed their understanding of the task, the experimenter launched the main version of the letter-*e* task, reiterated the instructions and left the room. The main test was exactly the same as the practice test, only it contained more trials and as a result took longer to complete. On returning to the room, the experimenter asked the participants to complete a brief 4-item questionnaire on effort, difficulty, fatigue and frustration experienced during the letter-*e* task. Instructions read: “Please respond to the statements below describing your feelings about the task you have just completed. **Circle one response** (one number) under each item”. The four questions and their accompanying scales included: “How much effort did you put in to the task? (1 = “No effort”, 7 = “A lot of effort”), How difficult did you find the task? (1 = “Very easy”, 7 = “Very difficult”), How tired do you feel after doing the task? (1 = “Not at all”, 7 = “Very much”), Did you feel frustrated while you were doing the task?” (1 = “Not at all”, 7 = “Very much”)”. Participants were informed that there were no correct or incorrect answers to these questions, and as such should be honest in their responses. The questionnaire items were answered on a 7-point scale and this measure of self-report was used in the manipulation checks to assess the effectiveness of the letter-*e* tasks at depleting ego.

The third and final test to complete was the 16 CCPJ trials. Two versions of the CCPJ task were created in E-Prime. They were identical, differing only in the order in which the descriptions and statements were presented. For example, in version 1, the Bill scenario was presented as a UL statement; however, in version 2 it was presented as an RI statement (see Appendix A for a full list of the scenarios and their corresponding statements). The two versions were counterbalanced between participants. This was done to control for any potential learning and order effects. The task introduction and instructions read: “People are not always given all the necessary information they need to make a good judgement. Sometimes we have to make the best possible judgement with the limited information we have. This study requires you to make 16 social judgements. You must first read a short personality description about a fictitious person, and then make a true/false judgement based on the limited information you received in the description. Let's begin by practicing a trial together." After the practice trial, the participants completed the 16 CCPJ trials on their own; however, the experimenter remained in the room and was available to answer any questions that arose over the duration of the study.

Each trial began with a fixation cross presented for 1 second followed immediately by the prime slide with instructions that read "Please read the following description carefully. You will not be able to return to this page." The short personality description was presented below these instructions and the participants were given unlimited time to read it. To proceed to the target slide they were prompted to press the spacebar. The target slide presented the statement sentence and asked the participants to answer if they thought the statement was correct. Participants pressed "A" for yes or "L" for no. After every answer, the participants saw a slide asking them to state how confident they felt that the answer they made was correct. They rated their confidence levels between 0 ("not at all confident") and 9 ("extremely confident"), using the keyboard to enter their answers. After completing 16 trials, participants saw a slide that signalled the end of the task, thanked them for participating and asked them to call the experimenter. The experimenter then paid the participants £4 each, issued them 30 sona credits if they had signed up to participate through the sona-system, and debriefed them before dismissing them.

Results

Careless responders and outliers. One hundred and forty-five participants were recruited for this study. Participants were screened for careless responding (Meade and Craig, 2012), where an overall "flag score" was computed to identify careless responders who were identified as outliers on a number of predetermined criteria. The first of which was the total duration spent completing the CCPJ task. A *z*-score was computed from each total duration time and any participants who fell outside of three standard deviations were flagged as outliers. There were no participants flagged on this criterion. Next, I screened for multivariate outliers on judgement latencies. Mahalanobis distance scores were computed for the length of time it took participants to think about their answers and make their judgements. Twenty-nine participants were flagged on this criterion. Following this, I screened for multivariate outliers on reading latencies. Mahalanobis distance scores were computed for the length of time it took participants to read the scenarios (personality vignettes) in the task. Seventeen participants were flagged on this criterion. Finally, a total flag score was computed for each participant using these three criteria and all participants who were flagged more than once were removed from the sample. Five participants were removed from the sample based on these grounds ($n_{control} = 1$, $n_{low\ depletion} = 2$, $n_{high\ depletion} = 2$), leaving the final sample at $N = 140$ ($n_{control} = 49$, $n_{low\ depletion} = 48$, $n_{high\ depletion} = 43$).

Manipulation Checks. Assuming that heuristic deliberations impede logical intuitions, an *e*-task was administered to deplete participants' cognitive resources resulting in the impairment of deliberative thinking and thus an increase of heuristic considerations. After completing either the high or low depletion version of the *e*-task, and in line with Sripada et al. (2014), all participants answered four self-report screening questions about their feelings regarding the task. Independent samples t-tests confirmed that participants who performed the high ego depletion task admitted putting more effort into the task than those who completed the low depletion version of the task, $t(90) = -4.32$, $p < .001$, Cohen's $d = 0.91$, see Table 4. 1 for descriptive statistics. They also found the task more difficult, $t(90) = -12.55$, $p < .001$, Cohen's $d = 2.62$, and became more frustrated during the task than those in the low depletion condition, $t(90) = -5.36$, $p < .001$, Cohen's $d = 1.12$. Surprisingly however, there was no self-reported difference in levels of tiredness after doing the two tasks, $t(90) = -0.28$, $p = .782$, Cohen's $d = 0.06$.

Table 4. 1

Mean self-report scores measuring how much effort participants put into the task, how difficult they found the task, how frustrated they felt during the task, and how tired they felt after completing the task, as a function of ego depletion condition. Scores were measured on a 7-point likert scale ranging from 1 (lowest) to 7 (highest)

	Low depletion (*n* = *49*)			High depletion (*n* = *43*)		
	M	*SD*	95% CI	*M*	*SD*	95% CI
Effort	4.29	1.86	[3.75, 4.82]	5.67	1.06	[5.35, 6.00]
Difficulty	1.76	0.95	[1.48, 2.03]	4.77	1.34	[4.34, 5.18]
Frustration	1.98	1.39	[1.58, 2.38]	3.65	1.60	[3.16, 4.14]
Tiredness	3.39	1.79	[2.87, 3.90]	3.49	1.67	[2.98, 4.00]

On completion of each CCPJ trial, participants were asked to rate how confident they were that their judgement was correct. This was done on a 10-point Likert scale (0 = "not confident at all", 9 = "extremely confident"). An average confidence score was computed for each participant across the sixteen CCPJ trials. An independent samples t-test revealed there was no difference in confidence ratings between the two depletion conditions, $M_{low\ depletion}$ = 6.91, SD = 1.27 vs. $M_{high\ depletion}$ = 6.74, SD = 1.00; $t(90) = 0.69$, $p = .493$, Cohen's d = 0.15. Next I compared confidence latencies (i.e., how long participants took to rate their confidence levels) across the two groups, and again found no difference, $M_{low\ depletion}$ = 2.04 seconds, SD = 0.68 vs. $M_{high\ depletion}$ = 1.95 seconds, SD = 0.72; $t(90) = 0.61$, $p = .545$, Cohen's d = 0.13. These findings suggest that any significant differences in the data are due to the manipulation of cognitive depletion and not due to differences in confidence scores or confidence latencies.

Conflict Sensitivity. To examine conflict sensitivity, a heuristic score was computed for each trial to assess whether participants were sensitive to the conflict between heuristic (i.e., representativeness) and logical considerations when making their judgements. See Experiment 1 for a detailed explanation on how these scores were computed. Heuristic scores ranged from 0% (never followed heuristic assessment) to 100% (always followed heuristic assessment) and were analysed using a 3 (depletion: control vs. low depletion vs. high depletion) x 2 (conflict: absent vs. present) mixed ANOVA with repeated measures on the last factor. Results revealed no significant main effect for conflict, $M_{no\_conflict}$ = 85.24, SD = 15.93, 95% CI [82.62, 87.86] vs. $M_{conflict}$ = 83.31, SD = 18.92, 95% CI [80.16, 86.45], $F(1,137) = 0.99$, $p = .323$, $\eta_p^2 = .01$. There was a significant main effect for cognitive depletion, $M_{control}$ = 88.52, SD = 11.51, 95% CI [84.87, 92.17] vs. $M_{low\ depletion}$ = 83.33, SD = 13.42, 95% CI [79.65, 87.02] vs. $M_{high\ depletion}$ = 80.96, SD = 13.84, 95% CI [77.07, 84.86], $F(2,137) = 4.18$, $p = .017$, $\eta_p^2 = .06$. The pairwise comparison of the heuristic score across load conditions showed that people in the control condition were significantly more heuristic in their judgements compared to those in the high depletion condition ($p < .05$); whereas there was no difference in heuristic scores of those in the control condition and low depletion condition ($p > .05$), or the low depletion condition and the high depletion condition ($p > .05$). There was no interaction between conflict and cognitive depletion, $F(2,137) = 0.60$, $p = .551$, $\eta_p^2 = .01$ (see Table 4. 2 for descriptive statistics). The effect of conflict was not larger under higher load; thus, I cannot conclude that load moderates the effect of conflict.

Table 4. 2

Means, Standard Deviations and 95% Confidence Intervals of Heuristic Scores for No Conflict and Conflict Statements across the Control, Low Depletion and High Depletion Conditions

	Control			Low depletion			Ego depletion		
	M	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
No conflict	90.05	12.49	[85.63, 94.48]	82.81	16.83	[78.34, 87.29]	82.85	17.47	[78.12, 87.57]
Conflict	86.99	17.30	[81.68, 92.30]	83.85	19.63	[78.49, 89.22]	79.07	19.43	[73.41, 84.73]

The lack of conflict detection in this data is perplexing as there is compelling evidence for conflict detection in both experiments 1 and 2. For this reason, I revisited the confidence ratings and confidence latencies (from the manipulation checks) to test whether sensitivity to the conflict between heuristic and logical considerations could be found. As currently any sign of conflict sensitivity is being masked by the cognitive depletion tasks. If conflict sensitivity arises out of the substitution process (i.e., substituting an easier action for a more difficult one) then in addition to reduced self-report confidence judgements, we should also expect to see longer confidence response latencies for the conflict statements. In other words, if participants are questioning whether their answer is justified, this uncertainty should translate into increased processing time on the conflict statements, relative to the no conflict statements (for which the immediate heuristic response is also the logical response). Additionally, if participants feel uncertain about their responses, it may take more time to translate this uncertainty into a precise estimate of confidence compared to when one is fully confident. Hence, measuring the time participants take to provide a confidence judgement might provide an additional index of conflict sensitivity.

Participants rated how confident they were that their judgement was correct on a 10-point Likert scale (0 = "not confident at all", 9 = "extremely confident"). These judgements were also timed to give the confidence latencies. A mean confidence rating score as well as a mean confidence latency score was computed for each participant across the conflict and no-conflict trials. Paired samples t-tests confirmed that participants were more confident in their judgments when conflict was absent from statements, $M_{no\ conflict}$ = 6.97, SD = 1.24 vs. $M_{conflict}$ = 6.69, SD = 1.23, t(91) = 3.01, p = .003, Cohen's d = 0.23; and participants took longer to rate their confidence when conflict was present in the statements, $M_{no\ conflict}$ = 1.93s, SD = 0.74 vs. $M_{conflict}$ = 2.02, SD = 0.78, t(91) = -2.20, p = .03, Cohen's d = 0.19. Conflict detection was not present in the CCPJ responses; however, in the confidence ratings and confidence latencies there was evidence that people are sensitive to the conflict between heuristic and logical considerations.

Judgement Latencies. Response latencies were analysed to assess which responses showed sensitivity to conflict between heuristic and logical considerations. Assuming that people who always respond with a heuristic judgement under conflict use a different strategy overall, we first screened and excluded them, so we could compare latencies for heuristic and logical answers under conflict (N = 97: $n_{control}$ = 23, $n_{low\ depletion}$ = 36, $n_{high\ depletion}$ = 38; 31% removed). In line with previous research (Pennycook, Fugelsang, & Koehler, 2012; De Neys & Glumicic, 2008) three average latency scores were computed for each participant: 1) the average response latency for congruent statements (R&L and U&I statements), 2) the average response latency for heuristic responses to conflict statements (R/I and U/L statements), 3) and the average response latency for logical responses to conflict statements. See Experiment 1 for more details on how to compute these three latency scores. Latencies were converted to $\log^{10}$ prior to analysis to normalize the distribution and subjected to a 3(judgment type: no-conflict vs. conflict-heuristic vs. conflict-logical) x 3(load: control vs. low depletion vs. high depletion) mixed ANOVA with repeated measures on the first factor.

As expected, there was a significant main effect for ego-depletion condition with latencies being faster in the two ego-depletion conditions than the control condition, $M_{control}$ = 4.34, SD = 0.20, 95% CI [4.25, 4.42] vs. $M_{low\ depletion}$ = 4.16, SD = 0.20, 95% CI [4.09, 4.23] vs. $M_{high\ depletion}$ = 4.16, SD = 0.20, 95% CI [4.09, 4.23], F(2, 80) = 6.62, p = .002, η_p^2= .14

(see Table 4. 3 for descriptive statistics in seconds). The pairwise comparisons showed a difference in latencies between the control and low depletion conditions ($p < .05$), and a difference between the control and high depletion conditions ($p < .05$). There was no difference between the low and high depletion conditions ($p > .05$). There was also a main effect of judgement type with participants taking longer to make judgements when conflict was present in the statements, and even longer to make logical judgements with conflict statements, $M_{no\_conflict} = 4.02$, $SD = 0.19$, 95% CI [3.97, 4.06] vs. $M_{conflict\_heuristic} = 4.07$, $SD = 0.18$, 95% CI [4.04, 4.11] vs. $M_{conflict\_logical} = 4.54$, $SD = 0.41$, 95% CI [4.47, 4.64], $F(2, 160) = 137.32$, $p < .001$, $\eta_p^2 = .63$ (see Table 4. 3 for descriptive statistics in seconds). The pairwise comparisons show a significant difference in judgement latencies between the no conflict and conflict-heuristic judgements ($p < .05$), the no conflict and conflict-logic judgements ($p < .05$), and the conflict-heuristic and conflict-logic judgements ($p < .05$). There was no interaction between load condition and judgement type, $F(4, 160) = 0.18$, $p = .947$, $\eta_p^2 = .01$, thus load condition did not moderate judgement latencies. See Figure 4. 3 for descriptive statistics.

Table 4. 3

Means, Standard Deviations and 95% Confidence Intervals as a function of Depletion Condition and Judgement Type. Data are expressed in seconds

		M	*SD*	95% CI
Depletion Condition	Control	31.60	18.78	[25.89, 37.30]
	Low depletion	22.04	10.42	[16.96, 27.12]
	High depletion	21.50	11.98	[16.59, 26.41]
Judgement type	No conflict	11.77	5.71	[10.67, 12.87]
	Conflict-heuristic	13.24	6.94	[11.84, 14.63]
	Conflict-logical	50.13	36.19	[42.24, 58.02]

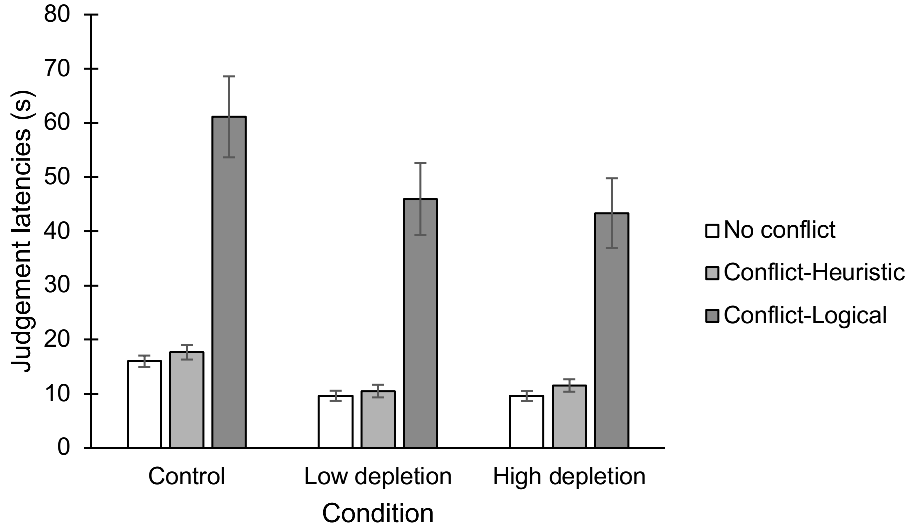


Figure 4. 3. Judgement latencies as a function of response type across the experimental conditions. Note. Data is presented in seconds and error bars represent standard error

Impulsivity and Logical Sensitivity. To further understand the origin of the conflict between heuristic and logical assessments, individual differences in impulsivity were evaluated to see whether they would predict logicality. BIS-11 scores were analysed to ascertain whether participants who were more impulsive were also more logical in their considerations when answering conjunction judgements. A person who scores high in impulsiveness should in theory be more likely to use heuristic thought processes when making conjunction judgements (Gerrard et al., 2008). If this is the case, then according to the hypothesis that heuristic considerations result in more logical conjunction judgements, those participants with higher BIS-11 scores should answer the CCPJ trials more logically than those with lower scores.

A logical sensitivity score was computed for each participant, which showed how reliant they were on logical considerations when answering the 8 conflict trials. One participant who scored 100% logicality was removed from the sample. Consistent with the conflict sensitivity analysis, higher cognitive depletion resulted in more logical sensitivity;

however, this difference was not significant, $M_{low\ depletion}$ = 16.00, *SD* = 19.40, 95% CI [11.00, 21.75], $M_{high\ depletion}$ = 20.00, *SD* =19.47, 95% CI [14.44, 26.11], $t(93) = -1.00$, $p = .319$, Cohen's $d = 0.21$. Responses to the BIS-11 ranged from 63 to 122 (low impulsivity – high impulsivity). To test whether logic sensitivity was predicted by impulsivity as a function of cognitive depletion, a regression model was run regressing a mean deviation form of BIS-11 scores, the depletion condition with the contrast code -1 = high depletion, 1 = low depletion, and the interaction between the BIS-11 scores and the load condition. The model did not provide a significant fit to the data, $F(3,91) = 1.28$, $p = .285$, $R^2 = .041$. Contrary to my expectations, the interaction term did not reach statistical significance ($p = .328$). Figure 4. 4 presents the scatterplot for the model and Table 4. 4 shows the detailed results of the regression analysis.

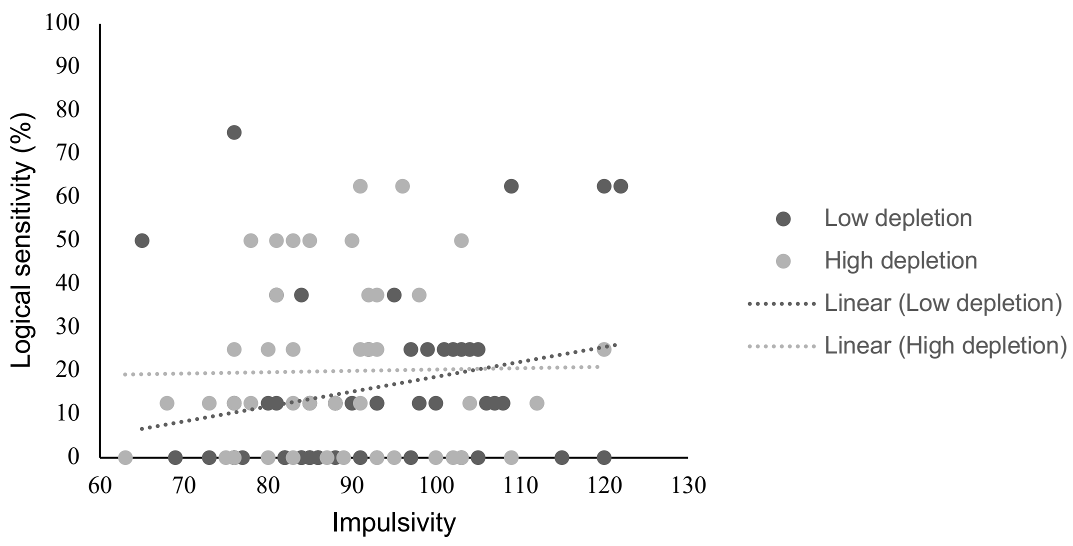


Figure 4. 4. Scatterplot of the interaction between BIS-11 scale and load condition on logical sensitivity

Table 4. 4

Linear regression analysis predicting sensitivity to the conflict between logical and heuristic assessments from depletion condition (-1 = high depletion, 1 = low depletion), and impulsivity

	B	95% CI	β	*p*
Constant	17.78	[13.80, 21.75]		< .001
BIS-11	0.19	[-0.13, 0.50]	0.12	.241
Load condition	-2.28	[-6.26, 1.70]	-0.12	.259
BIS-11 x Load condition	0.16	[-0.16, 0.47]	0.10	.328

Note. R^2 = .041, $F(3, 91) = 1.28$, p = .285

Discussion

Support of Hypotheses and Interpretation of Results. The objective of this study was to test the hypothesis that decreasing individuals' capacity for effortful thinking would also have a positive impact on probability judgments because it would obstruct the inhibition of the rapid, logical response by conscious effortful deliberation. As with the preceding experiments, it also hypothesised that individuals would show a sensitivity to the conflict underlying intuitive and logical considerations, and alongside this they would also demonstrate logical intuitions and heuristic reflections. To be more specific, it was expected that participants who were depleted cognitively would respond more logically as they were responding in line with intuitive processes. Whereas, those in the control condition, and possibly the low depletion condition would have access to their deliberative processing and thus commit more fallacies as the representativeness heuristic biases reflective thinking. It tested these hypotheses by identifying the impact of depletion of cognitive resources on the acceptance rates of both logical and illogical statements. The results found in this study partially supported the hypotheses. Evidence of conflict sensitivity was found in the confidence ratings – both in their confidence responses and their confidence response latencies – and also in their judgement latencies across types of responses. Individuals were more confident in their judgments when conflict was absent from statements and took longer to rate their confidence when conflict was present in the statements. Evidence of conflict sensitivity was also present in the judgement latencies data. Participants took significantly longer to make judgements when conflict was present in the statements, particularly when they judged the conflict statements logically (i.e., correctly).

Participants in the control condition were slower to complete the CCPJ task than those in the low depletion and high depletion conditions. There was however, no difference in time between the two depletion conditions ($M_{control}$ = 31.60 seconds vs. $M_{low\_depletion}$ = 22.04 seconds vs $M_{high\_depletion}$ = 21.50 seconds). This evidence ties in nicely with ego-depletion studies such as Baumeister et al. (1998) and Muraven et al. (1998) who found that volition expended ego, which resulted in less persistence on subsequent problems. In the case of this study, depleted participants might have given up quicker on the CCPJ task than the control group, thus resulting in their shorter judgement latencies. If this is in fact the case, then we might assume that in order to achieve these faster response times they relied more heavily on their faster, heuristic system of thought instead of their slower, more effortful deliberative thought. There was a significant effect of ego-depletion such that participants in the control group responded more heuristically (i.e., they answered more in line with heuristic considerations than logical considerations) than those in the high depletion group ($M_{control}$ = 88.52 vs. $M_{high\_depletion}$ = 80.96). On average, participants in the high depletion group made their judgements 10 seconds faster than those in the control condition. These results show that ego-depleted participants who performed the letter-*e* task with conditions, also made the most correct judgements in the shortest time. This lends support for the idea that heuristic responses can reflect logical considerations; however, when people take time to deliberate their judgement is biased by non-logical heuristics. However, the existence of logical intuitions and heuristic reflections should be researched further, as there is no way of being certain which system of thought the participants were using when they arrived at their answers. As mentioned in previous chapters, the methodology cannot predict if participants were guessing. It is possible that depleted participants were so cognitively drained that they just gave up and guessed their answers. Additionally, most of the significant differences found in the data of this experiment are due to the inclusion of the control condition from Experiment 1. The strength of the data presented here would have been higher if a new control had been collected for this experiment.

Finally, this study explored whether the propensity to be impulsive had any bearing on one's logicality when judging conjunction probability judgements. Due to the higher order factors of this scale, including self-control and perseverance, a positive correlation between those individuals who perceived themselves as non-impulsive and an increased propensity to think deliberatively was expected. In keeping with this study's hypothesis that the logical answer comes first, but is overridden by a heuristic bias, it was also expected that individuals who assessed themselves as being impulsive, also exhibited higher logicality in their conjunction judgements. Unfortunately, there was no evidence found to support these theories. Despite previous work that shows impulsivity reflects a person's propensity to deliberate less (Dickman, 1990), these findings suggest that a person's propensity to be impulsive has no relationship to the considerations they use when making conjunction probability judgements.

Generalizabitily of Results. Perhaps the most important question to arise from these results is if logical responses are in fact fast, then why did this not show in the judgement latencies data. These results showed that participants who answered the conflict statements logically did so in the slowest times. In fact, they were four times slower to answer conflict statements in line with logical considerations compared to when answering conflict statements in line with heuristic considerations, or when answering the no conflict statements (see Table 4. 3 and Figure 4. 3). Indeed, these findings are more closely aligned with the traditional two-system model whereby the slower, more effortful System 2 thinking is necessary for reaching logical conclusions. In fact, the parallel-competitive account explains these findings perfectly. Intuition and deliberation work in parallel from the beginning of the reasoning process allowing people to detect when conflict arises between the two systems of thought. Conflict detection triggers a deeper processing of the task at hand, which can result in the slower, more deliberate process inhibiting the faster, more intuitive process, thus resulting in a logical response. However, this is not always the case and sometimes (in fact most of the time when it concerns the conjunction fallacy) the heuristic response is too enticing or convincing to be overridden.

Looking closer at the judgement latencies across the response types, you will notice the enormous standard deviations for the conflict-logical judgements (SD = 36.19 seconds). This incongruity prompted me to explore the conflict-logical data in a bit more depth. A scatterplot of the data helps to illuminate the reasons for these conflicting findings (see Figure 4. 5). When observing only the conflict statements, it is noticeable that those participants who answered the fastest also answered the most statements logically. The faster they responded to the conflict statements, the more correct answers they gave. Conversely, participants who only gave one or two correct answers out of a possible eight, also took the longest to respond. These findings support the logical intuitions account, that faster judgements lead to higher logicality. In other words, the fast heuristic response can indeed be logical, while the slower more deliberative response can be biased by prior beliefs and heuristics. The slow, but logical responses can be explained by the idea that it takes a great amount of cognitive energy and deliberation to reinstate the logical answer after it has been overridden by a potent heuristic reflection. Thus, it is difficult, and takes time but not impossible to reinstate the original logical intuitive answer. This suggests that people can have both fast and slow reasoning processes that point towards the logical response. However, I also admit the necessity to research these ideas further, as it is always a possibility that the depleted participants were giving up quickly when conflict was present in the statements and guessing.

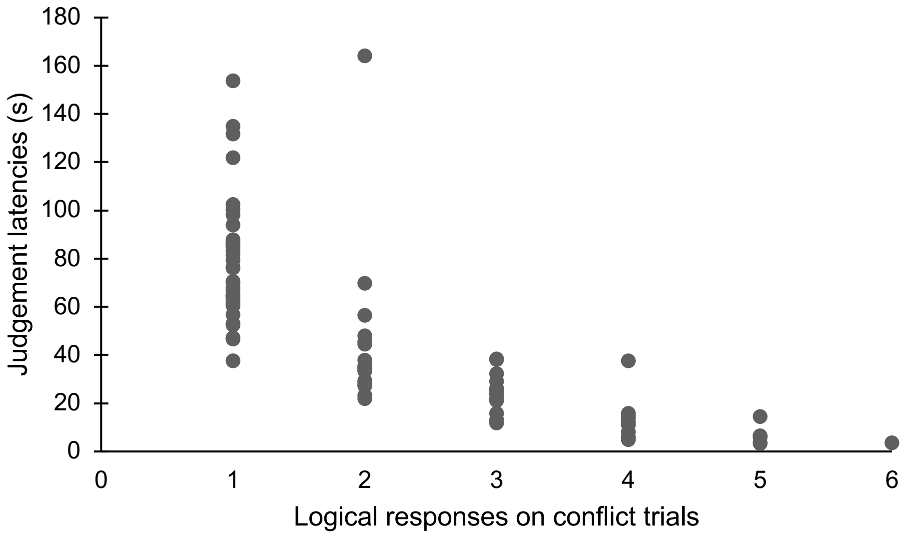


Figure 4. 5. Scatterplot of the time taken in seconds to respond logically to conflict trials

The manipulation between high and low depletion seemed to make no difference to the extent that participants were depleted. Why did the low depletion version of the letter *e* task not work in this study? This could be due to the actual task as described by Myers et al. (2018) who suggest that the letter *e* task with conditions is just a more difficult version of the letter *e* task without conditions. If this is the case, then both versions of the task might have had the same depleting effect on the participants, which might also explain the similarity in self-reported tiredness felt after completing the *e*-task in both the low and high ego-depletion conditions ($M_{low\ depletion}$ = 3.39, $M_{high\ depletion}$ = 3.49, p = .782). The task manipulations (i.e., crossing out all examples of the letter *e* compared with crossing out the letter *e* according to a set of rules) are designed to increase task difficulty, but this might not ultimately induce ego-depletion. Increasing the complexity and difficulty of the task would obviously result in poorer performance; however, task difficulty alone cannot account for declined performance over time (Myers et al., 2018). "A more difficult task will mean poorer performance (a main effect in a sense) when compared to a less difficult task but it won't necessarily produce greater deficits in performance over time" (Myers et al., 2018, p. 4). The fact that there was no difference in ego depletion between the low and high depletion condition, in a way makes total sense, as in line with the limited strength model people have a finite storage of energy that can become depleted when they undertake tasks requiring self-control. When this happens, people need to replenish these energy stores before continuing, as this will negatively affect performance on subsequent self-control tasks. In the manipulation checks, participants from each depletion condition completed a questionnaire pertaining the extent to which they felt, or experienced, depletion. These self-report answers did indeed differ between the two groups for all questions except how tired they felt after the task. Participants in the high depletion condition reported feeling more frustrated, putting more effort into the task and found the task more difficult than those in the low depletion condition. However, when it came to analysing their behavioural data there were no differences found between these two groups. What this might mean for our understanding of these energy stores is that once they are depleted, there is nothing left. There does not seem to be varying levels of depletion. In other words, one person cannot be more depleted than another. You either have energy reserves remaining or you do not. This would also have implications for ego-depleting tasks, such that a task is either depleting or it is not. You would not be able to find one task that is more depleting than another. However, this is purely speculative and would need to be tested with further research.

A further explanation as to why the ego depletion conditions did not differ from each other could be a methodological one. One of the major assumptions of the limited-strength model is that in order for ego depletion to occur, an act of self-control needs to be undertaken; in other words, a prepotent response needs to be overridden. “Self-control entails overriding or altering a predominant response tendency” (Schmeichel & Vohs, 2009, p. 770). Without a habit-forming section at the beginning of the task, there is no prepotent response to override. The experimental condition is supposed to consume more self-regulatory resources than the control, and the reason for this is straightforward. The instructions of the first task – to cross out every letter *e* – is designed to build and embed a habit. While the instructions of the second task – to NOT cross out the letter *e* in specific cases – forces the participants to inhibit the prepotent response. This inhibiting or overriding of the newly acquired habitual response is the cause of resource-depletion (Myers et al., 2018). Indeed, Baumeister and Vohs (2016) view the lack of a habit-forming stage prior to the instigation of a new set of rules as a methodological flaw. They believe that without the formation of a habit there is nothing to override, and this overriding of an impulse is essential to the nature of self-control tasks.

The presence of conflict sensitivity was not found in the heuristic scores in this experiment. Thus, one could interpret that the participants were not aware of the conflicting assumptions between logical and intuitive thinking. According to the reinforcement sensitivity theory the behavioural inhibition system (BIS) resolves conflicts between the behavioural activation system and the fight-flight-freeze system. The BIS is a conflict detection system that helps to balance one impulse in favour of another (Schmeichel, Harmon-Jones & Harmon-Jones, 2010). BIS has also been identified as playing a role in the detection of response conflict in cognitive tasks (e.g., the go/no go task). Amodio, Master, Yee and Taylor (2007) found that higher BIS reflected greater conflict-related anterior cingulate cortex (ACC) activity, and they concluded that BIS “corresponds to an attentional system for monitoring response conflicts” (p.16). Cognitive neuroscience believes that the ACC serves a conflict-monitoring function, whereby it monitors for conflict and recruits additional mechanisms to resolve such conflicts. Inzlicht and Gutsell (2007) found that depleting cognitive resources had an impact on performance on the Stroop test. Participants were required to either inhibit or express their responses to a distressing film clip. They found that exercising self-control (i.e., inhibiting a distressing response) reduced activity in the brain's conflict-detection centre, thus reducing BIS activation. This explanation might account for the reason why conflict detection was absent from the heuristic scores in the present study. If the letter *e* tasks (both the low and the high depletion conditions) caused ego-depletion in the participants, it is possible that this also resulted in the reduction of BIS activation, and henceforth a lack of conflict detection. This assumption; however, is not corroborated by the self-report confidence ratings and the response times taken to make these confidence ratings, as there was evidence found of conflict sensitivity in these. Participants were less confident in their judgements and took longer to rate their confidence when conflict was present in the statements.

Finally, there is a potential confound in the data, such that it is unclear whether the ego depletion manipulations did in fact deplete cognitive resources and induce heuristic thinking, or whether ego depletion simply compelled participants to answer the task faster, thus causing higher levels of logicality (as explained by the logical intuitionist account). Participants in both the low and high depletion conditions completed the CCPJ task significantly faster than those in the control condition. These faster response times might in fact be the reason why the participants were answering more in line with System 1 thinking. Both time pressure and ego depletion are recognised as measures that encourage System 1 thinking, thus the question remains, which one was responsible for the decrease in heuristic scores between the control and manipulation groups ($M_{\_control}$ = 88.52, $M_{\_low\ depletion}$ = 83.33, $M_{\_high\ depletion}$ = 80.96). Although, when re-examining the judgement latencies across the three conditions ($M_{\_control}$ = 31.60s, $M_{\_low\ depletion}$ = 22.04s, $M_{\_high\ depletion}$ = 21.50s) they do not appear to match the speed of those in the limited time pressure condition of Experiment 1

($M_{\_limited}$ = 4.38s), and neither do the heuristic scores ($M_{\_limited}$ = 76.72). Thus, the reason for decreasing heuristic scores as ego depletion increases is most likely a result of the ego depletion task itself, and not faster response times.

Final remarks and Future Directions. Although the ego depletion manipulation appeared to have been successful in this experiment, there are still doubts remaining. There is substantial evidence in the literature providing support for the fact that ego depletion and the limited strength model might not be as robust as previously believed (Moller et al., 2006; Tice et al., 2007; Job et al., 2010; Carter & McCullough, 2014; Hagger et al., 2016; Martijn et al., 2002). "Recent conceptual and empirical analyses have challenged the resource depletion explanation for the self-regulatory failures observed in ego-depletion experiments and questioned the strength of the ego depletion effect or whether it exists at all" (Hagger et al., 2016, p. 547). However, there was significant drop in heuristic scores from the control to the high depletion condition, which suggests that the manipulation might in fact have made a difference to people's logicality. The heuristic scores remained very high at 88.52% and 80.96% respectively, but the high depleted participants made almost 10% fewer errors than those in the control condition. Another question to consider is why there was no difference between the two depletion condition. The manipulation checks showed significant differences in self-reported ratings of effort, frustration and difficulty but there was no difference between the heuristic scores. This might suggest that the low and high depletion conditions in fact did not differ from one another, which in turn questions the reliability of the ego depletion task.

With this uncertainty in mind, future research might benefit from using a more developed methodology to collect the data. In other words, a task that incudes multiple answers to choose from instead of only two, which would eliminate any suspicions of guessing. It would also benefit from a more robust method of supressing cognitive resources whilst performing the CCPJ task. This prompted the next experiment, which utilises a cognitive load task to overload cognitive resources as a method of encouraging System 1 thinking. The task involves a secondary memory task to overload working memory while answering the CCPJ task. The thought behind this is that by overloading working memory during the task, one is encouraging intuitive thought as there will be less capacity to deliberate one's judgements prior to answering.

Experiment 4: The Influence of Spatial Storage Load and Cognitive Reflection on the Conjunction Fallacy

Dual-Process Theories and the Conjunction Fallacy. Dual processes have been at the core of trying to explain the shortcomings of human judgements (i.e., the failure of logic and rational thinking). Most of the time these two reasoning systems coexist harmoniously. This happens when the heuristic default system provides a fast, automatic and correct response that aligns with the logical response of the analytic system. This is an example of congruency, when there is no conflict present between logical and heuristic considerations. However, this is not always the case; sometimes the intuitive response is at odds with the deliberative response. In other words, they are each pointing to a different answer. When this happens, the deliberative system must override the default heuristic response to replace the first intuitive response with a logical one (Stanovich & West, 2000). "Participants who are higher in System 2 resources (i.e., cognitive ability) are more likely to suppress the incorrect response from System 1 in favour of System 2, and give the correct response" (Feeney, Shafto & Dunning, 2007, p. 885). Indeed, the authors found that participants who performed poorly on the AH4 (a widely used group test of cognitive ability) were more susceptible to committing the conjunction fallacy. This lends support for the two-system notion that one needs to be able to deliberate effectively in order to override the prepotent heuristic response and replace it with a logical one.

Working Memory and Judgement. There are studies consistent with the dual process framework which show that individual differences in cognitive capacity, such as working memory, can predict performance on conflict trials, but not on congruent trials (e.g., Stanovich & West, 2000; De Neys, 2006a). System 1 is assumed to operate automatically; hence it should not overburden limited executive resources. People with poorer working memory should be able to arrive at the correct answer for trials that do not contain conflict, as the default heuristic response aligns with the logical response. However, on trials containing conflict, only people with higher cognitive abilities will be able to override the heuristic response and replace it with a logically correct one. Crisp and Feeney (2009) found that a secondary memory task increased conjunction errors for strongly related events but had no effect on the weakly related scenarios. They suggest that under memory load, most participants lack the cognitive resources necessary to inhibit the compelling heuristic output activated by a strong causal link. However, when the causal link is weaker, the heuristic response is less compelling, thus participants can overcome it even under memory load. Finally, the authors suggest that "causal knowledge has a crucial influence on probabilistic reasoning" (Crisp & Feeney, 2009, p. 2331).

The Cognitive Reflection Test (CRT; Frederick, 2005) is a brief, three-item task that is designed to measure a person's ability to override an incorrect prepotent heuristic response and replace it with a logical and correct response through engaging in further reflection. The bat-and-ball problem is the best-known CRT item. In this problem, participants are asked:

"A bat and a ball together cost £1.10.

A bat costs £1 more than a ball.

How much does the ball cost?"

The appealing but incorrect answer is "10p". This answer is believed to be generated effortlessly and automatically by intuitive processes. The correct answer to this problem is in fact "5p" and requires that effortful deliberation be employed to inhibit and override the prepotent intuitive response. The CRT has become a popular measure of individual differences (see Appendix E for the full version of the CRT). A few reasons for this being that the CRT is a performance measure rather than a self-report measure, and neither intelligence tests nor measures of executive functioning assess the tendency toward miserly processing in the way that the CRT does (Toplak, West & Stanovich, 2011). It is also very short and easy to conduct both as a pen and paper test or electronically. Additionally, it has been shown to be able to discriminate well between impulsive and reflective decision makers (Oechssler, Roider, & Schmitz, 2009; Travers, Rolison & Feeney, 2016).

Although the CRT is a convincing predictor of intelligence and cognitive aptitude (Frederick, 2005; Obrecht, Chapman, and Gelman, 2009), it is also a unique and potent predictor of performance on heuristics-and-biases tasks (Toplak, West & Stanovich, 2011), and more pertinently on conjunction fallacies (Oechssler, Roider, & Schmitz, 2009). In this study, participants were presented with the classic "Linda task" by Kahneman and Tversky (1983) and asked to indicate which of these two statements was more likely to be true: 1) "Linda is a bank teller" or 2) "Linda is a bank teller and is active in the feminist movement". The authors found that higher test scores on the CRT (i.e., scoring 2 or more correctly out of 3) were correlated with lower incidences of the conjunction fallacy. However, even in the higher CRT group, committing the conjunction fallacy remained fairly established at 38.3% (compared to 62.2% for the low CRT group). Hoppe and Kusterer (2011) found that CRT test scores were able to predict performance on certain heuristics and biases tests. More specifically, they found that individuals with lower cognitive abilities, as measured by the CRT, are significantly more likely to exhibit both the base rate fallacy and the conservatism fallacy. Interestingly however, they found no link between CRT test scores and the endowment effect, which was striking in both low and high CRT groups. Toplak et al. (2011) argue that low CRT scores indicate a tendency to behave as a "cognitive miser", giving the first response which comes to mind. In other words, these decision makers "act on impulse" and neglect to employ deliberation to override the prepotent heuristic response (Alós-Ferrer & Hügelschäfer, 2012, p. 183).

The CRT and Dual Process Theories. Dual process theories differ in their account of CRT performance. The default-interventionist models (Evans, 2006; Kahneman & Frederick, 2002) propose that intuition is the default mode of processing and that deliberative processing needs to be actively engaged to override and inhibit the initial, fast intuitive response. However, this does not always happen and often, even after engaging deliberation, people fail to adequately replace the intuitive response with a logical one (Stanovich & West, 2008). Failure to engage deliberative processing has been used to explain incorrect responding (i.e., heuristic responding) on the CRT (Travers, Rolison, & Feeney, 2016). More specifically, when a heuristic answer is given, deliberative processing has likely failed to engage. However, when a correct answer is given, it is believed that the initial heuristic response has been inhibited and replaced by a logical one. In other words, system 1 processing is overridden by system 2 processing. Travers, Rolison and Feeney (2016) used a novel methodology incorporating mouse tracking to capture time-course of reasoning on the CRT. They found that participants were initially drawn towards the intuitive (i.e., incorrect) option even if they proceeded to choose the deliberative (i.e., correct) answer. Furthermore, participants who chose the incorrect answer, were not attracted to the correct option prior to this. They concluded that intuitive processes are automatically activated and must be inhibited to answer correctly. They add that when participants responded intuitively, there was no evidence that deliberation had taken place. These findings lend support for the default-interventionist models of thinking and performance on the CRT.

Parallel-competitive dual process theories posit that both heuristic and deliberate processing are activated simultaneously resulting sometimes in conflict that needs to be resolved. When this conflict happens, both systems compete for control of behaviour (Evans, 2007; Sloman, 1996). Similarly to default-interventionist models, these accounts predict that system 1 processing needs to be inhibited by system 2 processing in order to reason correctly. Distinctively though, parallel-competitive models predict that system 2 processing is persistently attempting to signal the correct response, even when failing to override the output of system 1 processing.

The Current Study. Dual process theories have come under heavy criticism due to the fact that the framework has focussed primarily on people's responses while ignoring the mechanisms underlying the cognitive processes (De Neys, 2006b; Reyna, Lloyd & Brainerd, 2003; Stanovich & West, 2000; Gigerenzer & Reiger, 1996). The tasks that are commonly used to assess the conjunction fallacy offer a narrow window on the mechanisms that may underpin judgments. This may result in confounds as studies that include and explore individual differences are purely correlational and thus showed no causality (De Neys, 2006a).

Using a new methodology adapted from Evans & Curtis-Holmes (2005) that disentangled logic from intuition, Villejoubert (2009) found that individuals were more likely to accept representative statements, but they were also significantly less likely to accept illogical ones. Moreover, they consistently took longer before accepting a statement when heuristic and logical considerations were in conflict, even when they were pressured to provide an answer under time pressure. The current and widely accepted understanding is that when the slow deliberative system is not given enough time to process information, thinking will be more heuristic and less logical. However, Villejoubert's (2009) findings suggest that quick thinking can be logical while effortful thinking can be biased by non-logical heuristics.

The present research aims to build on these ideas and explore more systematically the circumstances under which accurate logical considerations may be fast and intuitive in judgements of probability. The objective is to test original and theoretically-driven ways to improve individuals' probability judgements, and this will be addressed by testing the hypothesis that decreasing individuals' capacity for effortful thinking will also have a positive impact on probability judgments because it will impede the inhibition of the rapid, logical response by conscious effortful deliberation.

Using the Comparative Conjunction Probability Judgement (CCPJ) task methodology, this study aims to identify the impact of increased cognitive load on the acceptance rates of both logical and illogical statements. Increasing cognitive load has produced conflicting findings with standard conjunction tasks (Crisp & Feeney, 2009; De Neys, 2006a; Villejoubert, 2009). This research will provide an assessment of the impact of cognitive load on the relative weight of heuristic and logical considerations in probability judgments. Oechssler, et al. (2009) raised an interesting question in their research: are people who exhibit the conservation bias less susceptible to the base rate fallacy? In the same train of thought, one might also ask the question: are people who exhibit examples of System 1 thinking, also less susceptible to committing the conjunction fallacy in conjunction probability judgements? This study also aims to investigate whether a relationship exists between cognitive reflection and logicality on conjunction judgements. The hypothesis is that people who exhibit lower CRT scores will also commit fewer conjunction fallacies, as they will be less likely to inhibit the logical intuitive answer. People with higher CRT scores are more able to inhibit intuition and employ deliberation and thus will commit more conjunction fallacies as their deliberation will be biased by heuristic assessments.

Method

Participants. Power analysis performed in GPower (Faul, Erdfelder, Lang & Buchner, 2007) indicated that a total sample size of 159 students would be sufficient to detect effects of a medium size effect (d = .5), as reported in Villejoubert (2009), with 1 - b = .80 and a = .05. One hundred participants were recruited for this study. A control group of 50 participants was added to this study from the unlimited condition in Experiment 1. In this condition, participants simply completed the 16 CCPJ trials in their own time and capacity. They were not exposed to any time pressure nor were they put under any cognitive load. This brought the total sample size to 150 participants. There were two participants who responded logically to all of the eight conflict trials, in other words, these participants were unbiased in their answers, thus were removed from the sample to avoid a possible confound in the data leaving the final sample size at 148 students. Following screening for careless responding (details outlined in results section), a further 7 participants were ruled out leaving the sample at 141 students ($n_{control} = 48$, $n_{low\ load} = 46$, $n_{high\ load} = 47$). Participants ranged in age from 18-56 years ($M = 23.70$, $SD = 7.19$). One hundred and eighteen were females (83.7%) and 23 males (16.3%). Most of the students were undergraduates ($n = 84$, 59.6%) and the remaining 57 students were postgraduates (40.4%). One hundred and five of the students were native English speakers (74.5%), while 24 rated themselves as intermediate English speakers (17%), and the remaining 12 as proficient in English (8.5%).

Materials

The Comparative Conjunction Probability Judgement (CCPJ) task. This is a series of 16 CCPJ trials presented on a computer screen. Each trial presented a short personality vignette describing an imaginary person's hobbies and characteristics, followed by one of four types of probability statements: representative and logical (RL), representative but illogical (RI), unrepresentative but logical (UL) and unrepresentative and illogical (UI). Participants saw four of each type of statement presented in a random order, they then had to judge whether they believed the statement to be true or false, and entered their answers using the keyboard of the computer. See Chapter 2 for full details of the CCPJ task.

Dot memory task. This classic spatial storage task can originally be seen in the work presented by Bethell-Fox and Shepard (1988) and Miyake and colleagues (2001). It is well-established that it successfully burdens participant's executive resources (De Neys & Schaeken, 2007; Franssens & De Neys, 2009; Johnson, Tubau, & De Neys, 2016; Miyake et al., 2001). For this study, a 3x3 matrix filled with three or four dots (depending on which load condition: low or high) was briefly presented to each participant for 850 ms (see Figure 4. 6). Participants memorized a new pattern before reading each social judgement vignette and were asked to reproduce it after each judgement they made. In the recall phase, an empty matrix was presented on the screen, and participants used the mouse to indicate the location of the dots. Hovering over an empty square and left clicking the mouse would cause a dot to appear in the corresponding location. If the participant decided to change his/her mind, hovering over the dot and left clicking the mouse a second time would clear the square.

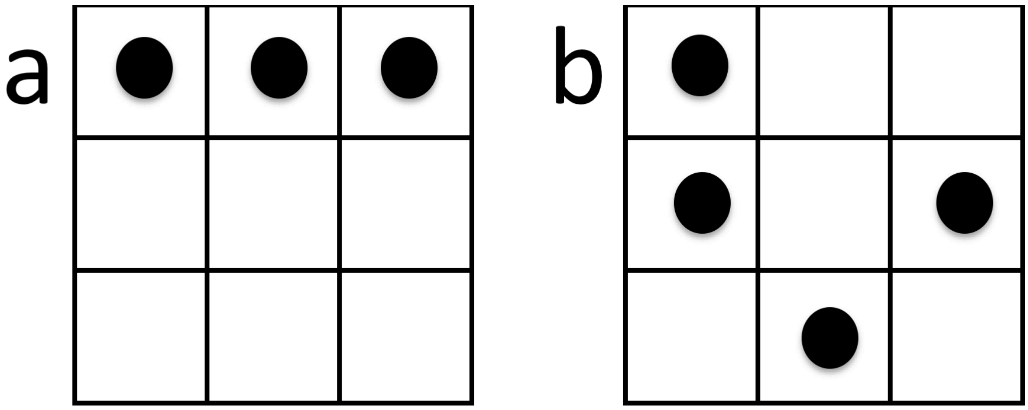


Figure 4. 6. Examples of the dot patterns in the low-load (a) and high-load (b) conditions

In the high load trials, the matrix was filled with a complex four-dot pattern (i.e., a "two- or three-piece" pattern based on the work of Bethell-Fox & Shepard, 1988, and Verschueren, Schaeken, & d'Ydewalle, 2004). Miyake et al. (2001) established that memorising complex dot patterns such as these depleted executive resources. In the low load trials the pattern consisted of three dots on a horizontal, vertical or diagonal line (i.e., a "one-piece" pattern in Bethell-Fox & Shepard's terms). This simple and systematic pattern should only minimally burden the executive resources (e.g., De Neys, 2006a; Miyake et al., 2001).

Cognitive Reflection Task (CRT). Taken from Frederick (2005), this test is composed of three questions, all of which are designed to produce an intuitive response. In other words, they all hint towards a fast heuristic response in the first instance. This fast response is also the incorrect response, and in order to replace it with the correct answer one must override or suppress this first response and employ reflective analytical thinking (Frederick, 2005). See Appendix E for the full CRT. Frederick (2005) accompanied his CRT with the following prior instructions: “Below are several problems that vary in difficulty. Try to answer as many as you can.” For this experiment I replaced these instructions with the simple heading of “Brain teaser”. A combined measure of performance on these three items was the dependent variable (i.e., a score out of 3).

Procedure. In this experiment, participants made conjunction probability judgements while they simultaneously tried to remember a briefly presented visual dot pattern (e.g., Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; De Neys, 2006a). The difficulty of the dot pattern was manipulated so that remembering the pattern in a control condition would be easier and less demanding on executive resources. As a measure of individual differences, they also completed the CRT (Frederick, 2005). On arrival the participants received an information sheet and consent form. After being briefed on the study and giving their informed consent, they commenced with two tasks: the CRT and the CCPJ trials. The order in which they completed these tasks was counterbalanced. The CRT was presented as a pen and paper task, while the CCPJ task was programmed in E-Prime 2 and presented digitally on a computer screen. The sequence of events to complete the CCPJ trials consisted of first reading a personality vignette about a fictitious person. Once they had finished reading this description, they pushed the spacebar and a dot matrix pattern flashed up on the screen for 850ms. After 850ms, the matrix disappeared from the screen and they had to read and answer the judgement statement. After making their judgement by entering their answer (“A” for True and “L” for False) on the keyboard, an empty matrix appeared on the screen and they had to recall the locations of the dots they had seen directly previously. The participants all completed three practice trials on the CCPJ task before they commenced with the full 16 experimental trials. They were instructed to focus predominantly on remembering the dot sequences. They were told that recalling the dots correctly was the most important part of the study and that their answers to the social judgements came secondary. This was to encourage participants to take the dot task seriously and ensure the experimental manipulation (i.e., overloading cognitive resources) succeeded. On completion of both the CCPJ task and the CRT, they were debriefed and paid in their choice of either course credits (30 credits) or money (£4). The study took less than 30 minutes to complete.

Results

Careless responders and outliers. Participants were screened for carless responding (Meade and Craig, 2012), where an overall "flag score" was computed to identify careless responders who were identified as outliers on a number of predetermined criteria. The first of which was the total duration spent completing the CCPJ task. A *z*-score was computed from each total duration time and any participants who fell outside of three standard deviations were flagged as outliers. There were no participants flagged on this criterion. Next, I screened for multivariate outliers on judgement latencies. Mahalanobis distance scores were computed for the length of time it took participants to think about their answers and make their judgements. Twenty-one participants were flagged on this criterion. Following this, I screened for multivariate outliers on reading latencies. Mahalanobis distance scores were computed for the length of time it took participants to read the scenarios (personality vignettes) in the task. Twenty-two participants were flagged on this criterion. A total flag score was computed for each participant using these three criteria (total duration, judgement duration and reading duration) and all participants who were flagged more than once were removed from the sample. Seven participants were removed from the sample based on these grounds ($n_{control} = 2$, $n_{low\ load} = 3$, $n_{high\ load} = 2$). Thus, 9 participants in total were removed from the sample as 2 responded 100% logically, leaving the final sample size at $N = 141$.

Manipulation Checks. Following De Neys and Schaeken (2007) and Johnson, Tubau and De Neys (2016), recall performance was assessed to ensure the dot memory task was properly performed. Largely, recall performance was high with the mean number of correctly recalled dots for the complex four-dot patterns reaching 3.48 (SD = 0.35) and 2.78 (SD = 0.22) for the simple three-dot patterns. In other words, 87% of a complex and 92.7% of a simple pattern was reproduced correctly, showing that the dot memory task was taken seriously by the participants and performed correctly.

Manipulation checks showed that there was a significant difference in total duration latencies (i.e., how long each participants took to complete the entire task from beginning to end) between load groups ($M_{control}$ = 478.19 seconds, SD = 163.05 vs. $M_{low\ load}$ = 718.96 seconds, SD = 263.88 vs. $M_{high\ load}$ = 732.72 seconds, SD = 305.33; $F(2,138) = 15.50$, $p < .001$). Pairwise comparisons revealed that total duration latencies differed between the control and low load condition ($p < .001$), and between the control and high load condition ($p < .001$), and there was no difference between the low load and high load conditions ($p = 1.00$). This was expected as the control condition did not fill out the dot pattern on a grid, nor were they required to rate their confidence after each trial, and thus were significantly faster than the other two conditions. The fact that there was no difference between the low and high load conditions is encouraging, as it means any differences in judgements between the two conditions can be attributed to the manipulation of cognitive load and not extra thinking time.

In the low load and high load conditions, on completion of each trial, participants were asked to rate how confident they were that their judgement was correct. This was done on a 10-point Likert scale (0 = "not confident at all", 9 = "extremely confident"). An average confidence score was computed for each participant across the conflict conditions of the CCPJ task. In other words, one score for trials containing conflict and another score for trials without conflict. These scores were subjected to a 2 (load: low vs. high) x 2 (conflict: absent vs. present) mixed ANOVA with repeated measures on the last factor. Results revealed a significant main effect of conflict. Participants reported feeling less confident in their answers when conflict was present in the statements ($M_{no\text{-}conflict}$ = 7.13, *SD* = 1.44, 95% CI [6.84, 7.43] vs. $M_{conflict}$ = 6.85, *SD* = 1.35, 95% CI [6.58, 7.13], $F(1,94) = 8.56$, $p = .004$, $\eta_p^2 = .08$). There was no main effect of load condition. Participants reported the same level of confidence in their answers across both load conditions ($M_{low\ load}$ = 7.01, *SD* = 1.26, 95% CI [6.63, 7.39] vs. $M_{high\ load}$ = 6.98, *SD* = 1.37, 95% CI [6.60, 7.35], $F(1,94) = 0.01$, $p = .909$, $\eta_p^2 = .00$). Finally, there was no interaction between conflict and load conditions ($F(1,94) = 0.76$, $p = .387$, $\eta_p^2 = .01$, see Table 4. 5 for descriptive statistics).

Table 4. 5

Means, Standard Deviations, and 95% Confidence Intervals across the load (low vs. high) and conflict (absent vs. present) conditions for confidence ratings (score out of 10)

				95% Confidence Interval	
Condition	Conflict	*M*	*SD*	Lower Bound	Upper Bound
Low Load	Absent	7.19	1.34	6.77	7.61
	Present	6.82	1.31	6.43	7.22
High Load	Absent	7.08	1.53	6.67	7.49
	Present	6.88	1.40	6.49	7.26

In addition to recording the confidence judgements, so were the latencies of how long it took each participant to make these judgements of confidence. These confidence latencies were subjected to a 2 (load: low vs. high) x 2 (conflict: absent vs. present) mixed ANOVA with repeated measures on the last factor. Results revealed a significant main effect of conflict. Participants took longer to rate their confidence when conflict was present in the statements ($M_{no\text{-}conflict}$ = 2758.76 ms, *SD* = 1127.70, 95% CI [2530.27, 2987.25] vs. $M_{conflict}$ = 2947.88 ms, *SD* = 1351.67, 95% CI [2674.00, 3221.75], $F(1,94) = 3.96$, $p = .049$, $\eta_p^2 = .04$). There was no effect of load condition. Participants took the same length of time to report their confidence across the two load conditions ($M_{low\ load}$ = 2713.96 ms, *SD* = 985.61, 95% CI [2379.72, 3048.20] vs. $M_{high\ load}$ = 2986.99 ms, *SD* = 1295.10, 95% CI [2659.65, 3314.34], $F(1,94) = 1.34$, $p = .249$, $\eta_p^2 = .01$). There was no interaction between the conflict and load conditions ($F(1,94) = 0.22$, $p = .639$, $\eta_p^2 = .00$, see Table 4. 6 for descriptive statistics).

Table 4. 6

Means, Standard Deviations, and 95% Confidence Intervals across the load (low vs. high) and conflict (absent vs. present) conditions for confidence latencies (in milliseconds)

				95% Confidence Interval	
Condition	Conflict	*M*	*SD*	Lower Bound	Upper Bound
Low Load	Absent	2642.09	1099.75	2315.46	2968.72
	Present	2785.82	1015.33	2395.03	3176.62
High Load	Absent	2870.67	1154.01	2550.77	3190.56
	Present	3103.32	1605.50	2720.58	3486.05

Conflict Sensitivity. To examine conflict sensitivity, a heuristic score was computed for each trial to assess whether participants were sensitive to the conflict between heuristic (i.e., representativeness) and logical considerations when making their judgements. See

Experiment 1 for a detailed description on how these scores were computed. Heuristic scores ranged from 0% (never followed heuristic assessment) to 100% (always followed heuristic assessment) and were analysed using a 3 (load: control vs. low vs. high) x 2 (conflict: absent vs. present) ANOVA with repeated measures on the last factor.

The results revealed that participants were more likely to respond heuristically in the absence of conflict between heuristic and logical considerations, $M_{no\_conflict}$ = 86.68, *SD* = 14.50, 95% CI [84.29, 89.07] vs. $M_{conflict}$ = 80.99, *SD* = 18.87, 95% CI [77.94, 84.04], $F(1,138) = 11.51$, $p = .001$, $\eta_p^2 = .08$. Heuristic thinking was also influenced by load condition; participants who were under a high cognitive load while making a judgement were least likely to rely on heuristic considerations, while participants under no cognitive load were most likely to rely on heuristic considerations, $M_{control}$ = 88.28, *SD* = 11.50, 95% CI [84.54, 92.02] vs. $M_{low\ load}$ = 84.38, *SD* = 12.69, 95% CI [80.56, 88.19] vs. $M_{high\ load}$ =78.86, *SD* = 14.89, 95% CI [75.08, 82.63], $F(2,138) = 6.20$, $p = .003$, $\eta_p^2 = .08$). Pairwise comparisons revealed that heuristic scores differed between the control and high load condition only (p = .002). There was no difference in scores between the control and low load condition (p = 452), nor between the low load and high load conditions ($p = .132$). I expected the low load condition to be similar to the control condition as it was designed to be very relaxed on the use of cognitive resources. However, the fact that there was no difference between the low and high load conditions is unexpected. Finally, there was no interaction between conflict and cognitive load, $F(2,138) = 0.98$, $p = .378$, $\eta_p^2 = .01$ (see Table 4. 7 for descriptive statistics).

Table 4. 7

Means, Standard Deviations and 95% Confidence Intervals of heuristic scores for the absence or presence of conflict as a function of load condition

Load	Conflict	*M*	*SD*	95% Confidence Interval	
				Lower Bound	Upper Bound
Control	Absent	89.84	12.54	85.75	93.94
	Present	86.72	17.38	81.49	91.95
Low	Absent	86.96	13.93	82.77	91.14
	Present	81.79	16.18	76.46	87.13
High	Absent	83.25	16.33	79.11	87.38
	Present	74.47	21.01	69.19	79.75

This result was surprising as an interaction was expected, assuming that cognitive load would not affect judgements in the absence of conflict between heuristic and logical considerations. However, it appears that cognitive load reduced both the proportion of heuristico-logical answers in the absence of conflict and the proportion of heuristic answers in the presence of conflict. The reasons for the impact of cognitive load on non-conflict scenarios are unclear. A possible reason might be that under load, people find it harder to process the unrepresentative statements (U&I, U/L). To test this idea, I ran a 3 (load: control vs. low vs high) x 2 (conflict absent vs. present) x 2 (statement: representative vs. unrepresentative) mixed ANOVA with repeated measures on the last two factors.

The results for the main effects of load, conflict and the interaction between load and conflict remained the same as outlined in the analysis above. In addition to this, there was a significant main effect of statement. Participants responded more heuristically to unrepresentative statements, $M_{representative}$ = 79.65, SD = 19.96, 95% CI [76.36, 82.95] vs. $M_{unrepresentative}$ = 88.02, SD = 14.70, 95% CI [85.69, 90.36], $F(1,138) = 20.20$, $p < .001$, η_p^2= .13. Finally, there was no interaction between statement and load ($F(2,138) = 1.50$, $p = .228$, η_p^2= .02), no interaction between statement and conflict ($F(1,138) = 0.02$, $p = .904$, η_p^2= .00), and no three-way interaction between statement and conflict and load ($F(2,138) = 0.58$, p = .561, η_p^2= .01). Thus, load did not moderate how difficult participants found the representative and unrepresentative statements. In general, and regardless of load, participants answered the representative statements more logically, while relying more heavily on heuristic considerations when making unrepresentative judgements. In line with Tversky and Kahneman's (1983) belief that cognitive misers rely on heuristics to make difficult judgements, one could assume that people find it more difficult in general to process the unrepresentative statements, regardless of cognitive load.

Judgement Latencies. Response latencies were analysed to assess which responses showed sensitivity to conflict between heuristic and logical considerations. Assuming that people who always respond with a heuristic judgement under conflict use a different strategy overall, they were screened for and excluded in order to compare latencies for heuristic and logical answers under conflict. The remaining sample was as follows: N = 83; $n_{\_control}$ = 23, $n_{\_low\ load}$ = 29, $n_{\_high\ load}$ = 31 (41% removed from sample). In line with previous research (Pennycook, Fugelsang, & Koehler, 2012; De Neys & Glumicic, 2008) three average latency scores were computed for each participant: 1) the average response latency for congruent statements (R&L and U&I statements), 2) the average response latency for heuristic responses to conflict statements (R/I and U/L statements), 3) and the average response latency for logical responses to conflict statements. See Experiment 1 for a more comprehensive explanation on how to compute these three latency scores. Latencies were converted to $\log^{10}$ prior to analysis to normalize the distribution and subjected to a 3 (judgment type: no-conflict vs. conflict-heuristic vs. conflict-logical) x 3 (load condition: control vs. low vs. high) mixed ANOVA with repeated measures on the first factor.

There was a significant main effect for judgement type. Participants were fastest at making judgements that contained no conflict, but took the longest amount of time to make logical judgements when conflict was present in the statements ($M_{no\_conflict}$ = 4.11, SD = 0.22, 95% CI [4.06, 4.15] vs. $M_{conflict\_heuristic}$ = 4.16, SD = 0.19, 95% CI [4.12, 4.20] vs. $M_{conflict\_logical}$ = 4.63, SD = 0.42, 95% CI [4.55, 4.72], $F(2, 188) = 154.47$, $p < .001$, $\eta_p^2 = .62$). There was also a significant main effect of load condition. Participants were overall fastest at making judgements when put under high load ($M_{control}$ = 4.34, SD = 0.20, 95% CI [4.24, 4.43] vs. $M_{low\ load}$ = 4.35, SD = 0.15, 95% CI [4.28, 4.43] vs. $M_{high\ load}$ = 4.21, SD = 0.29, 95% CI [4.14, 4.28], $F(2, 94) = 4.27$, $p = .017$, $\eta_p^2 = .08$). There was a significant interaction between load condition and judgement type. Participants were fastest at making judgements under high cognitive load, especially when conflict was absent from the statements, $F(4, 188) = 2.37$, $p = .054$, $\eta_p^2 = .05$ (see Figure 4. 7).

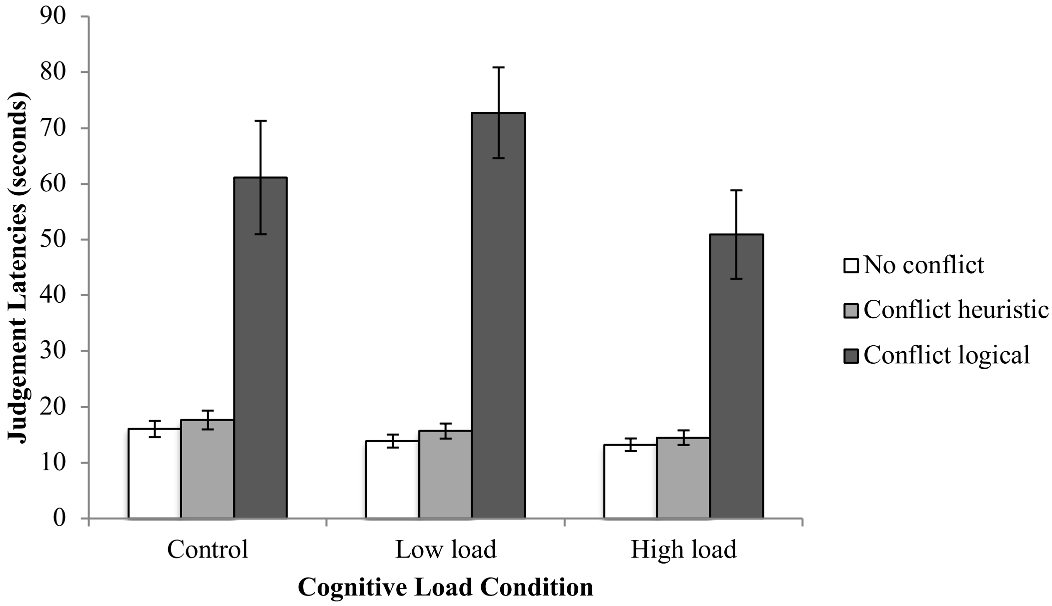


Figure 4. 7. Judgement latencies as a function of judgement type and load condition. *Note*. Error bars represent standard errors. The per-cell sample size was n = 23 for control, n = 36 for low load, and n = 38 for high load

Planned paired samples t-tests revealed that no-conflict statements were answered significantly faster than conflict-heuristic statements, $M_{no\_conflict}$ = 4.10, SD = 0.22, $M_{conflict\_heuristic}$ = 4.15, SD = 0.19, $t(96) = -3.55$, $p = .001$, Cohen's d = 0.24. No-conflict statements were also answered significantly faster than conflict-logical statements, $M_{no\_conflict}$ = 4.10, SD = 0.22, $M_{conflict\_logical}$ = 4.63, SD = 0.42, $t(96) = -13.42$, $p < .001$, Cohen's d = 1.58. Finally, conflict-heuristic statements were answered significantly faster than conflict-logical statements, $M_{conflict\_heuristic}$ = 4.15, SD = 0.19, $M_{conflict\_logical}$ = 4.63, SD = 0.42, $t(96) = -12.47$, $p < .001$, Cohen's d = 1.47. Please note, these descriptive statistics and paired samples t-tests are expressed in $\log^{10}$.

Cognitive Reflection and Logical Sensitivity. To further understand the origin of the conflict between heuristic and logical assessments, individual differences in cognitive reflection (i.e., the ability to override a heuristic response and replace it with a logical one) were evaluated to see whether they would predict logicality. If, as suspected, sensitivity to logic is fast and intuitive, people who are less cognitively reflective should be more logical in their judgements on the CCPJ task. A logical sensitivity score was computed for each participant. More specifically, one score that showed what percentage of time they judged conflict statements logically (0% = “never logical”, 100% = “always logical”). The sample was screened for perfect logicality and participants who always answered logically on the conflict trials were excluded from the analysis. Consistent with the conflict sensitivity analysis, higher cognitive load led to an increase in logical sensitivity and this difference was marginally significant, $M_{low\ load}$ = 17.82%, SD = 16.22, 95% CI [13.56, 22.34], $M_{high\ load}$ = 24.74%, SD =20.96, 95% CI [19.14, 30.61], $t(94) = 1.81$, $p = .074$, Cohen’s $d = 0.37$. Responses to the CRT ranged from 0 (never answered a trial logically) to 3 (answered all 3 trials logically), $M_{low\ load}$ = 0.45, SD = 0.88, 95% CI [0.21, 0.70], $M_{high\ load}$ = 0.37, SD = 0.67, 95% CI [0.20, 0.57]. To test whether logic sensitivity was predicted by cognitive reflection as a function of cognitive load, a regression model was run regressing a mean deviation form of CRT scores, the load condition with the contrast code -1 = high load, 1 = low load, and the interaction between the CRT scores and the load condition. The model did not provide a significant fit to the data, $F(3,92) = 1.45$, $p = .233$, $R^2 = .045$. Contrary to my expectations, the interaction term did not reach statistical significance ($p = .372$). Figure 4. 8 presents the scatterplot for the model and Table 4. 8 shows the detailed results of the regression analysis.

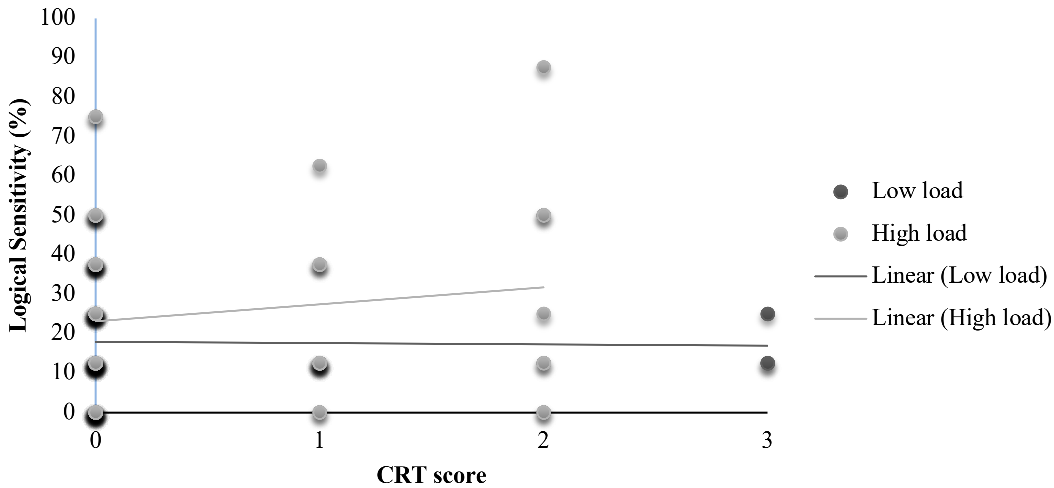


Figure 4. 8. Scatterplot of the interaction between CRT scores and cognitive load condition on logical sensitivity

Table 4. 8

Linear regression analysis predicting sensitivity to the conflict between logical and heuristic assessments from load condition (-1 = high load, 1 = low load), and cognitive reflection

	B	95% CI	ß	*p*
Constant	21.37	[17.54, 25.21]		< .001
CRT	1.99	[-3.14, 7.19]	0.08	.443
Load condition	-3.54	[-7.37, 0.29]	-0.19	.070
CRT x Load condition	-2.32	[-7.45, 2.81]	-0.09	.372

Note. R^2 = .045, $F(3, 92) = 1.45$, p = .233

Discussion

Support of Hypotheses and Interpretation of Results. This experiment examined the impact of cognitive load on the sensitivity to the conflict between heuristic and logical assessment in social judgements under uncertainty. The aim of this study was to test original and theoretically-driven ways to improve individuals' probability judgements. The results supported the hypothesis that people are able to detect the underlying conflict between the two systems of thought. The results also appeared to support the hypothesis that decreased capacity for effortful thinking would have a positive impact on probability judgments; however, the reason for this is still debatable. One explanation for the higher logicality observed in the depleted and high-load conditions could be that people might have access to logical intuitions. Indeed, this can be seen in the data as decreasing cognitive resources would have increased intuitive responses, which in turn resulted in higher levels of logical responses. The control conditions did not impose restrictions on the participants reasoning and allowed them to use conscious effortful deliberation. However, deliberation did not appear to equate with higher logicality (i.e., as predicted by the default interventionist and parallel competitive accounts) as these participants responded more in line with heuristic considerations; which seems to lend support for the heuristic reflections account. In other words, belief-based responses (i.e., responses derived from heuristics) are available as a System 2 response and depend on the inhibition of a competing intuitive-logical response (Howarth, Handley & Walsh, 2016). However, one must not ignore the fact that participants could also be guessing when confronted with the conflict trials as this conflict brings along uncertainty and the task requires a forced response. Due to the rate of heuristic responding being so high overall, it would only take a small amount of correct, or logical, guesses to sway the data.

The hypothesis that people who exhibited lower CRT scores would also commit fewer conjunction fallacies, as they are less likely to inhibit the logical intuitive answer was not supported. There was no significant relationship found between CRT sores and logicality. Oechssler et al. (2009) found that biases were significantly more pronounced for individuals with low cognitive abilities, as rated on the CRT. However, these results do not support this evidence. A reason for this could be that like making judgements in line with logical considerations, answering the CRT trials correctly is difficult and infrequent. Seventy-four percent of the participants never managed to get any CRT questions correct (see Table 4. 9 for the frequencies and percentages of the CRT scores). This could be due to a limitation in the procedure. The CRT was counterbalanced when being administered to the participants. Half answered it before completing the CCPJ task and half answered it afterwards. Those participants who answered it after the CCPJ task would have been submitted to cognitive load in the form of a dot test (either low load or high load), which might have fatigued them and affected their ability to solve the difficult CRT questions.

Table 4. 9

Frequency and percentages of CRT scores as a function of load condition

CRT score	Low load (n = 47)		High load (n = 49)	
	Frequency	%	Frequency	%
0	35	74.5	36	73.5
1	6	12.8	8	16.3
2	3	6.4	5	10.2
3	3	6.4	0	0

Generalizability of Results. It was apparent from both the heuristic scores and judgement latencies that the participants were able to detect the conflict underlying logical and heuristic considerations. Individuals answered more in line with heuristic considerations when conflict was absent from statements, especially in the control condition. In other words, the presence of conflict coupled with an increase in cognitive load saw an increase in logicality in individuals' answers. Participants also responded the fastest when conflict was absent from the statements, and there were significant differences in judgement latencies between the no-conflict, conflict-heuristic and conflict-logical judgement types. Such that the time taken to respond increased significantly between the judgement types in that order. This provides support for the notion that people are sensitive to conflict between logical and heuristic considerations (De Neys, 2012, 2014; De Neys & Schaeken, 2007; Villejoubert, 2009).

There was additional support for conflict detection found in the self-reported confidence ratings (e.g., how confident they were in their judgements) as well as the time it took participants to make these ratings (e.g., how long it took to assess their confidence levels). They were less confident, and also took longer to judge their confidence when conflict was present in the statements. When conflict was absent; however, they were more confident they had answered correctly and arrived at these confidence judgements faster. This suggests that people use a parallel processing model of reasoning as these models propose people are very good at detecting conflict.

There was no difference in load condition for both the confidence judgements and confidence judgement latencies. In other words, being exposed to higher or lower cognitive loads did not affect the participants belief in how accurately they had answered, nor did it affect the length of time they spent considering their accuracy. This result was surprising as the more complex 4-dot pattern was far more difficult to memorise and recall than the easier

3-dot pattern and I was expecting this to influence the confidence judgements. However, load condition had no effect on either the confidence ratings nor the confidence latencies. Indeed, the difference in load can be seen in the participant's heuristic scores: as the load increased, so did their logicality. There is evidence that people exhibit lower confidence when their response conflicts with the normative response (De Neys & Franssens, 2009; De Neys, Cromheeke & Osman, 2011). This literature has shown that even though individuals are unable to resist giving the intuitive response, they are also more sensitive to normative standards than is originally assumed as portrayed in their confidence scores. This goes against the default interventionist account and aligns itself more with a parallel model of dual processes. Stupple, Ball & Ellis (2013) found that the harder the problem, the lower the confidence levels. This was not supported in the current experiment as the high load task was far more difficult than the low load task but there was no difference in confidence scores. Thompson, Prowse, Turner and Pennycook (2011) found reduced 'feelings of rightness', which is closely associated with confidence, when participants *rejected* belief bias conclusions compared to when they *accepted* conclusions. Thus, it may be the case that people feel more confident when responding 'yes' to a statement, than responding 'no'. Indeed, Stupple et al. (2013) found that accepting, or agreeing with, a statement resulted in greater confidence than rejecting, or disagreeing with, a statement.

To investigate this idea further I computed two mean confidence scores for each participant for the conflict statements only. One for when they accepted statements and the other for when they rejected statements. A 2 (accept confidence vs. reject confidence) x 2 (low load vs high load) mixed ANOVA with repeated measures on the first factor was conducted on the data. There was a significant effect of type of response (i.e., accept / reject) such that individuals who rejected statements, actually reported feeling more confident than when they accepted statements (M_{accept} = 3.17, SD = 1.23, 95% CI [2.91, 3.43] vs. M_{reject} = 3.67, SD = 1.28, 95% CI [3.40, 3.93], $F(1,91) = 4.87$, $p = .03$, $\eta_p^2 = .05$). There was no difference in confidence levels between conditions ($M_{low\ \mathrm{load}}$ = 3.37, SD = 0.64, 95% CI [3.17, 3.57] vs. $M_{high\ \mathrm{load}}$ = 3.47, SD = 0.68, 95% CI [3.28, 3.64], $F(1,91) = 0.49$, $p = .488$, $\eta_p^2 = .01$) and there was no interaction between type of response and condition ($F(1,91) = .29$, $p = .592$, $\eta_p^2 = .00$). These findings do not support those found by Stupple et al. (2013), in fact they suggest that people feel more confidence in their answers when they reject or disagree with a statement.

Evidence for logical intuitions was also found in this experiment. Participants in the high load condition would also have had the least access to cognitive resources for deliberation, as their working memory was put under high load conditions (i.e., remembering a complex 4-dot pattern). The task was taken seriously by the participants with an accuracy score of 87% achieved, thus, we can assume participants in the high load condition were relying more heavily on their intuition while making their conjunction judgements. These participants responded significantly more logically (i.e., committed fewer conjunction fallacies) than those in the control condition. This evidence provides support that intuitive thought might point to the logical answer, even when there is conflict in the statements. The fact that the control condition had full cognitive ability, but still made more errors lends support for the heuristic reflections account. While it appears to be true that people can intuit logic, it might also be true that the bias is found in deliberative reflections, and it corrupts people's judgements when they analyse, or think hard about, their answers. The parallel-competitive account could explain the overall high levels of heuristic judgements despite the experimental manipulations. Whereby conflict is readily detected by people, but this does not necessarily translate into logical judgements due to the difficulty of inhibiting the potent heuristic response. Although, the data shows that participants with reduced cognitive capacity were in fact doing just this. How is it possible that people with less working memory capacity are able to perform a task that requires a lot of cognitive effort (e.g., inhibiting a potent response), while people who have ample cognitive capacity were not able to do so? One answer would be that the logical answer in fact comes first, and those people under cognitive load were in fact answering in accordance with their fast, automatic, intuitive – yet logical – reasoning process. However, it is important to note that the methodology employed to collect this data was a forced-choice task containing only two choices, which does not allow for the detection of guessing. It is true that people find judgements under uncertainty (e.g., trials containing conflict) difficult, thus they might resort to guessing under these conditions. Alternatively, they might simply be rejecting the heuristic response, thus choosing the only other alternative answer available to them. Or, they may in fact be reasoning logically. However, the methodology employed does not allow a differentiation between these possibilities, thus it is difficult to ascertain the reasons behind the increase of logical responses on the conflict trials and in the experimental manipulation conditions such as ego depletion and cognitive load.

Final remarks and Future Directions. To readdress the question raised at the beginning of this experiment: are people who exhibit examples of System 1 thinking, also less susceptible to committing the conjunction fallacy in conjunction probability judgements? From an initial examination of these data it would appear the answer is yes, although one must seriously consider how valid the data is given the methodology employed to collect it. Across this program of research there has been substantial evidence supporting the idea that people can detect conflict between the different systems of reasoning. There also appears to be evidence that people can intuit logic in conjunction judgement problems, and that heuristics can bias their analytical processing. However, these assumptions would need to be tested further using an expanded methodology that was better suited for pin pointing the reasons why people respond more logically when faced with conflict or when forced to respond intuitively. The question that is raised now, is how transferable this information is. Can it be applied to different heuristics and biases tasks? This question prompted the following and final experiment in this thesis, which explored whether the findings across the previous four experiments were transferable across different tasks and fallacies.

Chapter 5: Analysing Response Times to Further Understand Base Rate Neglect

Human judgement is not perfect especially under circumstances of uncertainty. A classic example of the unescapable impact that intuitive reasoning has on people's judgement and decision making is base rate neglect (Kahneman & Tversky, 1973). A typical base rate problem includes two different types of information. First, it includes base rate information in the form of background data that always includes numerical information about the composition of a sample. Second, it includes indicant or diagnostic information. This information is purely descriptive and evokes a stereotypical response. The base rate fallacy is when individuals tend to ignore base rates in favour of individuating information (when such is available) instead of integrating the two (Bar-Hillel, 1980). One exhibits the base rate fallacy when one allows indicators to dominate base rates when making probability judgements. Similar to the conjunction fallacy, people respond to these problems inappropriately because of the salient stereotypical response that is cued by the representativeness heuristic. Indeed, Kahneman and Tversky (1973) observed that the vast majority of educated university students and even university professors failed to answer the problem correctly. Their subjects applied their knowledge of the prior (i.e., base rate information) only when they were given no specific evidence (e.g., null description). "As entailed by the representativeness hypothesis, prior probabilities were largely ignored when individuating information was made available" (Kahneman & Tversky, 1973, p. 242).

Experiment 5: The Effects of Time Pressure and Cognitive Reflection on Base Rate Neglect.

Although it is clear that human judgement is often biased, the nature of the bias is unclear and misunderstood. It is commonly accepted that base rate neglect occurs because of the deliberative and effortful thinking of System 2 required to process base rate information, while fast and intuitive System 1 processing is required for diagnostic information (Kahneman & Frederick, 2002). These biased judgements are presumed to occur because the representativeness heuristic cues an intuitive response that is difficult to override (Kahneman & Frederick, 2002; Kahneman & Tversky, 1973). Additionally, some dual process theorists agree that a certain amount of deliberative System 2 processing is required for prior probabilities to enter into judgement, thus the difference of ease that base rates and stereotypes are processed is the source of base rate neglect (Bonner & Newell, 2010; De Neys & Glumicic, 2008; Kahneman & Frederick, 2002). This perspective is firmly grounded in the dual-process theoretical framework, which claims that human reasoning relies on two distinct systems of thought. System 1 is a fast, autonomous, intuitive system, while System 2 is slower, effortful and working memory dependent (Evans, 2008; Evans & Stanovich, 2013; Kahneman, 2003; Kahneman, 2011; Sloman, 1996).

Dual-process theory and base rate neglect.

The *default-interventionist* view suggests that people reason heuristically by default and that for the majority of the time they are unaware that these intuitions might be wrong (Kahneman, 2002; Evans, 2003). It is believed that due to lax monitoring people fail to detect that their intuitive response in fact conflicts with the laws of probability, thus “the dominance of intuitive thinking is attributed to a failure to monitor the output of the heuristic reasoning process” (De Neys, Vartanian & Goel, 2008, p.483). According to the *parallel-competitive* view, people always engage in probabilistic thinking and are thus able to detect when their intuitive responses are incorrect (Epstein, 1994; Sloman, 1996). Both systems of thought are engaged simultaneously from the beginning and operate in parallel. This enables people to detect a conflict between their stereotypical intuition and the appropriate response. However, the presence of conflict detection does not always translate into their responses. People are not always able to inhibit the prepotent intuitive response and thus "behave against their better judgement" (Denes-Raj & Epstein, 1994, p. 819). Thus, base rate neglect is attributed to an inhibition failure rather than to a lack of conflict detection.

According to the serial view (default-interventionist) a reasoner initially relies on the intuitive system and the deliberate system will only be cued if there is a conflict between the intuitive and deliberate outputs. There is a fundamental conceptual problem; however, as how will the reasoner ever detect the conflict between these two outputs if the deliberate system is not engaged yet? The parallel view resolves this problem as both systems are simultaneously activated; however, this view is also problematic as the deliberative system is engaged from the beginning, thus basically throwing away any benefit of an intuitive route (De Neys, 2012). When there is no conflict present (i.e., congruent trials), it is conducive to rely on the intuitive route. Engaging in demanding deliberation is superfluous in this situation and would be a complete waste of limited cognitive resources (De Neys & Glumicic, 2008). "What dual process models need is a way to detect whether deliberate thinking is required without having to engage in deliberate thinking" (De Neys, 2012, p. 34). The *logical intuitionist* account (De Neys, 2012) aims to solve this conceptual puzzle. It suggests that people have logical intuitions. In other words, they have an implicit knowledge of logical principles that is activated automatically. Thus, when faced with a reasoning problem people have two cued intuitive responses: one heuristic and one logical. When these responses conflict with one another, a noticeable arousal is created, which results in a questioning of the heuristic response. This experience will not be purely explicit, but more of an awareness that there is something wrong with the heuristic response; hence, conflict detection as a “gut feeling” (De Neys et al., 2010). The idea of a logical intuition has implications for dual process theories as it resolves the short comings of the serial and parallel views. “If the intuitive System 1 cues both a logical and heuristic response, potential conflict can be detected without prior engagement of System 2” (De Neys, 2013, p.8). If logical intuition is consistent with heuristic intuition, then people will select the cued response without any further deliberation. However, conflict between the two intuitive responses would signal the need to engage the deliberative system. It is important to note that even if deliberation is engaged, this does not imply a logical response as the heuristic intuition is very compelling. However, this theory does supply a switch rule that determines whether deliberation is required without engaging the deliberative system.

Conflict Detection in Base Rate problems.

The heuristics and biases paradigm has introduced many classic examples of erroneous human judgement and decision-making, including base rate neglect, also known as base rate fallacy or base rate bias. A classic example of the prevalent impact that intuitive heuristics have on human judgement and decision-making can be found in base rate neglect studies conducted by Kahneman and Tversky (1973). These studies required people to read a stereotypical description and respond to problems that cued a salient but inappropriate response. Participants first read a description about the composition of a sample of fictitious people (e.g., a sample of 70 lawyers and 30 engineers), they then read a short personality description of a randomly selected individual from that sample and had to indicate which group they thought the individual belonged to. Statistically, a randomly drawn individual would be drawn from the larger group, rather than the smaller group. However, due to the stereotypical beliefs cued by the personality description, people may be inclined to answer in line with the representativeness heuristic. For example: “Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles". Kahneman and Tversky (1973) found that the majority of their participants committed errors in judgement due to neglecting the prior base rate information and focusing on the stereotypical information at hand. To be more specific, most participants rated Jack as an engineer because of his personality description and ignored the fact that there were more than twice the number of lawyers than engineers present in the sample. "The failure to appreciate the relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction" (Kahneman & Tversky, 1973, p. 243).

An adaptation of the classical base rate task can be found in the following study. Consider the following problem taken from De Neys and Glumicic (2008):

In a study 1000 people were tested. Among the participants there were 995 nurses and 5 doctors. Paul is a randomly chosen participant of this study.

Paul is 34 years old. He lives in a beautiful home in a posh suburb. He is well spoken and very interested in politics. He invests a lot of time in his career.

What is most likely?

(a) Paul is a nurse.

(b) Paul is a doctor.

This problem is an example of an incongruent problem type. The personality description suggests that the answer is *doctor*, while the base rate information suggests it is *nurse*. When faced with this conflict, many studies have shown that people prefer to base their judgements on the individuating information found in the personality description and neglect or underweight the base rate information resulting in a base rate fallacy (for a review, see Barbey & Sloman, 2007).

In the last decade, there has been an increasing accumulation of evidence supporting conflict detection (De Neys & Glumicic, 2008; Thompson & Johnson, 2014; Pennycook, Fugelsang & Koehler, 2015; Villejoubert, 2009). In other words, it has become accepted that people who offer an intuitive response that conflicts with the principles of logic can often detect that their response is incorrect. To test for this sensitivity to conflict, researchers usually compare individuals' processing of the control and conflict versions of various judgement and decision-making tasks. Where control tasks (also known as congruent tasks) contain no conflict, but conflict tasks contain conflict between the cued heuristic response and the logical answer in line with statistical laws of probability. There is an assumption that if an individual is able to detect the conflict underlying the logical and heuristic considerations, this will also be evident in their processing. Indeed, these processing effects have been shown across a wide variety of tasks: biased reasoners who solve conflict versions typically need more time to reach their answers (Bonner & Newell, 2010; De Neys & Glumicic, 2008; Villejoubert, 2009; Pennycook, Trippas, Handley, & Thompson, 2014), they express less confidence about the correctness of their answers (Bago & De Neys, 2017; De Neys, Cromheeke, & Osman, 2011; Thompson & Johnson, 2014), and they show increased activation in their Anterior Cingulate Cortex (ACC) known to mediate conflict and monitor error (Ne Neys et al., 2008; Vartanian et al., 2018), compared to when they solve control or congruent versions.

In 2008, De Neys and Glumicic designed two experiments to establish how precisely people experience conflict between logical and heuristic considerations. The first experiment used a think aloud protocol alongside base rate neglect problems modelled after Kahneman and Tversky's (1973) classic problems. The second used problem processing times as an attempt to highlight the processes underlying these base rate neglect problems. Findings showed almost no evidence for conflict detection in the think aloud results; however, there was evidence of conflict detection in the recall data (i.e., a recall test completed after the base rate neglect problems to ascertain how deep the information was processed). This suggests that people might not be explicitly reporting an active struggle between the logical and heuristic considerations, but their cognitive system does seem to be detecting it on some level. Experiment 2 reinforced the results of Experiment 1 by showing superior recall on conflict and neutral trials. It also showed that better recall was accompanied by longer decision-making times and a tendency to review the base rate information more extensively. These results lend support the idea that a flawless conflict monitoring process exists, as whenever the base rates conflict with the description people are able to detect this and "consequently redirect attention towards a deeper processing of the base rate information" (De Neys & Glumicic, p. 1274).

With this in mind, Franssens and De Neys (2009) designed an experiment to test the automaticity of conflict detection. They hypothesized that the conflict monitoring process used minimal cognitive resources and operated quite automatically. The researchers had half their participants solve base rate neglect tasks under cognitive load (i.e., a difficult 4-dot memory test), while the remainder of the sample performed the task under no load. Afterwards they had to recall information regarding the base rates to highlight the depth at which this information was processed. Results replicated those of De Neys and Glumicic (2008) by showing the successful nature of the conflict-monitoring process. Recall was higher for those base rate problems that contained conflict than for those without conflict. More interestingly however, was the recall performance of those participants who had to make judgements under load. Cognitive load had no effect on the conflict monitoring process under load, which suggests that it requires only minimal resources to operate. In the load condition, base rate recall remained better for conflict trials than for no-conflict trials. Additionally, recall performance on conflict trials were equally high in the load condition as in the no load condition. In other words, the load condition impeded the deliberative thinking needed to solve the base rate problem, but it did not prevent conflict detection from occurring. This evidence supports the notion that the conflict monitoring process is both successful and effortless. Research has shown that people do in fact detect conflict between heuristic and logical considerations even though this does not always translate into their answers. De Neys, Vartanian and Goel (2008) found that the inhibition area of the brain (lateral prefrontal cortex) was activated when stereotypical responses were avoided, but more importantly they also found that the conflict-detection area (anterior cingulate) was activated even when people reasoned stereotypically. These findings suggest that people are detecting their bias even when they respond intuitively.

Aims and rationale

Understanding and clarifying the exact nature of the heuristic bias is important for the development of judgement and decision-making theories. However, it is difficult to decide between these existing views as the majority of the studies focus on the accuracy of the output (i.e., whether or not people give the correct response to the problem), and not on the underlying processes (De Neys et al., 2008). The current and widely accepted understanding is that people are "cognitive misers", thus when the slow deliberative system is not given enough time to process information (e.g., forced to answer quickly, or intuitively), thinking will be more heuristic and less logical (Gilovich, Griffin, & Kahneman, 2002). However, recently there have been a number of studies suggesting that intuitive thinking can be logical (De Neys, 2012; Villejoubert, 2009; Handley & Trippas, 2015; Villejoubert, 2009). Indeed, fast thinking can in fact be logical while slow thinking can be biased (Howarth, Handley & Walsh, 2016; 2018). The previous four experiments in this program of research attempt to demonstrate logical intuitions. They also attempt to offer an explanation as to why there are such high rates of probability errors in the control conditions (i.e., even when people are allowed to think hard about their answers). The default interventionist model has been ruled out due to the fact that people are very capable of detecting the conflict underlying intuitive and deliberative processes, providing support instead for a parallel model of thinking.

Recently a new theory has emerged by Howarth, Handley and Walsh (2016, 2018). It supports the notion that human reasoning operates as two parallel process competing with each other. The authors agree with the idea of logical intuitions and propose that the logical response is present from the beginning of the reasoning process, but it is often inhibited by heuristic reflections. In other words, effortful thinking (e.g., deliberation or reflection) can be biased by heuristics (e.g., prior beliefs or heuristics). If this is in fact true, then it might help to explain the high rates of heuristic responding across the program of research thus far. As discovered across the previous four experiments, people are not mere cognitive misers. They are sensitive to the conflict between competing responses, and they actually wanted to deliberate and think hard about their answers as can be seen in their response times from the control condition (e.g., they spent 19 seconds on average per trial). However, this observation also lends support for the parallel-competitive account that states the representative answer is very potent and difficult to suppress, but this account believes that the representative response results from fast System 1 thought. So these two different accounts need to be addressed and tested further.

The purpose of this study is to build on the novel idea of logical intuitions by confirming that individuals are sensitive to logical principles when using System 1, intuitive reasoning. The study will use base rate and time pressure manipulations in an attempt to identify the cognitive underpinnings of probability judgements and will test the hypothesis that "individuals will judge conflict statements (i.e., inconsistent with base rate considerations) as more logical than congruent statements (i.e., consistent with base rate considerations), and that under time pressure individuals' will answer more problems correctly, or logically". This is because people have an unlearned, inherent understanding of the logical principals that underlie probability judgements, which is represented in their intuitive answers. It will also test the heuristic reflections account that proposes when people are given time to think about their answers, their deliberation can be biased by non-logical heuristics.

Using an adaptation of the Comparison Conjunction Probability Judgement (CCPJ) task used in the previous experiments, this study aims to investigate the effect that conflict (between logical and heuristic considerations) and time pressure have on the use of the representativeness heuristic in base rate judgements. This study aims to replicate the findings of Experiment 1 by assessing the effect of conflict on heuristic and logical answers, plus it will use severe time pressure as a method of inducing System 1 reasoning to investigate whether individuals are sensitive to conflict between the two systems of reasoning, as well as to assess what effect increased System 1 reasoning has on base rate errors. Furthermore, the study aims to assess if cognitive reflection has any impact on an individuals' logicality when making base rate judgements. Following the results in Villejoubert (2009), as well as the results from Experiment 1, I hypothesize that both conflict and time pressure will increase logicality in base rate judgements. There should be convincing evidence, in both rate of heuristic responding as well as response latencies, for the fact that people are able to detect the conflict underlying logical and heuristic statements, as well as the observation that people under time pressure make fewer base rate errors, as they should be relying on their intuition to make these judgements. Finally, I expect to see that individuals who showed little or no cognitive reflection on the Cognitive Reflection Test (CRT; Fredericks, 2005), to also have answered more logically. Cognitive reflection is the ability to override an intuitive response and replace it with a deliberative one. If my assumption is correct, and our intuition can reflect logical principles, but deliberation causes prior beliefs to bias our answers, then being able to respond intuitively should produce more logical judgements. In other words, people who are unable to inhibit the intuitive response should also show more logicality in their responses to the base rate task.

Method

Participants. Power analysis indicated that a sample of 34 participants per group (i.e., 68 participants in total as there were two groups) would be sufficient to detect effects of a medium size effect (d = .5), as reported in Villejoubert (2009), with 1 - b = .80 and a = .05. The final sample consisted of 76 Kingston University students ranging in age from 18 – 57 years (M = 21.60, SD = 6.32). Sixty-four were females (84.2%) and 12 males (15.8%). Most of the students were undergraduates (n = 71, 93%) while the remaining five were postgraduates (7%). All the participants were psychology students. The majority were majoring in Psychology (n = 64, 84%); however, there were some who had second majors including Biology (2%), Creative Writing (2%), Criminology (4%), Drama (2%), English (2%), French (2%), and Sociology (2%).

Procedure and Materials. Participants were tested individually in a laboratory room on the Kingston University campus. They all completed the base rate task electronically on the laboratory computer as well as the Cognitive Reflection Task (CRT; Frederick, 2005), which was completed as a pen and paper test. The base rate task and the CRT were administered in a random order. The CRT was printed on a single page under the heading "Brainteaser". The participants were instructed verbally to complete as many of the questions as possible. See Appendix E for an example of the three-item CRT. The CCPJ task from the previous experiments was adapted for this study. It contained the same 16 vignettes; however, instead of being a conjunction fallacy task it was tailored to be a base rate neglect task. This was done by adding base rate information to each scenario. To be more specific, the base rate information was displayed before the personality description (see Figure 5.1). For example, "Imagine a sample of 150 people consisting of 5 accountants and 145 rock musicians. The description below was chosen at random from the 150 available." This was then followed by a description of Bill (e.g., an accountant as described in the original task by Tversky and Kahneman, 1983). Participants were then instructed to press the space bar following which the statement appeared for them to either accept as true or reject as false (e.g., “Bill is more likely to be an accountant than a rock musician”). This statement was answered either under time pressure, or no time pressure, depending on the experimental condition the participants were allocated to. They logged their answers by pressing “A” for YES, or “L” for NO on the keyboard. Finally, they logged their confidence in their response on a 10-point scale ranging from “0” = not at all confident to “9” = extremely confident.

Conflict versus no conflict was operationalised by the wording of the statements combined with the information provided by the base rates. The statements were either congruent with the base rate information provided at the start of the trial (i.e., the statement contained no conflict), or they were incongruent with the base rate information (i.e., the statement contained conflict). The trial type (e.g., R&L, R/I, U/L, and U&I) was designed by manipulating the large vs. small populations in the base rate information with the statements. For example, in a situation where rock musicians outnumbered accountants (as worded above), judging the following statement as true (e.g., “Bill is more likely to be an accountant than he is to be a rock musician”) would be making a representative but illogical (e.g., R/I) judgement because the participant answered in accordance with the personality description and ignored the base rate information. If the participant had judged the statement as false, they would have made an unrepresentative and logical (e.g., U/L) judgement because they used the base rate information and disregarded the personality description.

On arrival at the experiment, all participants were first given the information sheet to read, any questions they had were answered by the researcher, and finally when they were satisfied, they signed the consent form. The researcher read through the instructions with the participant before starting the base rate task, and there was one practice example to complete in each time condition before the 8 trials began (participants completed 8 trials under time pressure and 8 trials without time pressure). The participants had opportunities to ask questions during the example trials and when they felt ready they could begin the task. The order of time conditions was randomised and counterbalanced. Half of the participants completed the first 8 trials without any time pressure followed by the second 8 trials under time pressure. The other half completed the first 8 trials under time pressure, followed by 8 trials with no time pressure.

No time pressure instructions included the following instructions, “Please take all the time you need to read the following description” and “Please carefully consider all the information available before reaching a decision”. Whereas time pressure instructions included “You must give your first ‘gut’ response as you will not have long to think about your answers. Each time you take too long to answer you will lose credits, so remember – answer fast!” and “Please HURRY and record the first answer that comes to your mind!” Time pressure was encouraged through the loss of credits. Each participant began the test with 8 credits (each credit was worth 50p, which equated to £4), and each time they took too long to provide a judgement they would lose a credit, thus losing money. They were allocated 7500ms to read the statement and make their judgement. This amount of time allowed the participants approximately 2.5 seconds to read the statement and 5 seconds to make a judgement. This was calculated on an average reading speed of 350 words per minute. Adults read at a speed of 300 wpm but this can rise to as much as 400 wpm for university students, and each statement was between 12 – 16 words in length. Taking longer than 7.5 seconds resulted in the loss of a credit accompanied by a message displayed in red text on the screen that read, “Careful! You took too long to respond. Remember to respond as fast as you can to retain your credits. You currently have [number] credits remaining. You have lost a total of [loss] credits.” If the participants answered faster than 7500ms, they received a message in green text that read, “Good! You responded fast enough. Continue to respond as fast as possible to retain your credits. You have [number] credits remaining.” See Figure 5. 1 for examples of a trial containing no time pressure and a trial containing time pressure. Following every judgement they made, participants had to rate how confident they felt about their judgement being correct. They rated this on a scale from 0-9, “0” = not at all confident, and “9” = extremely confident.

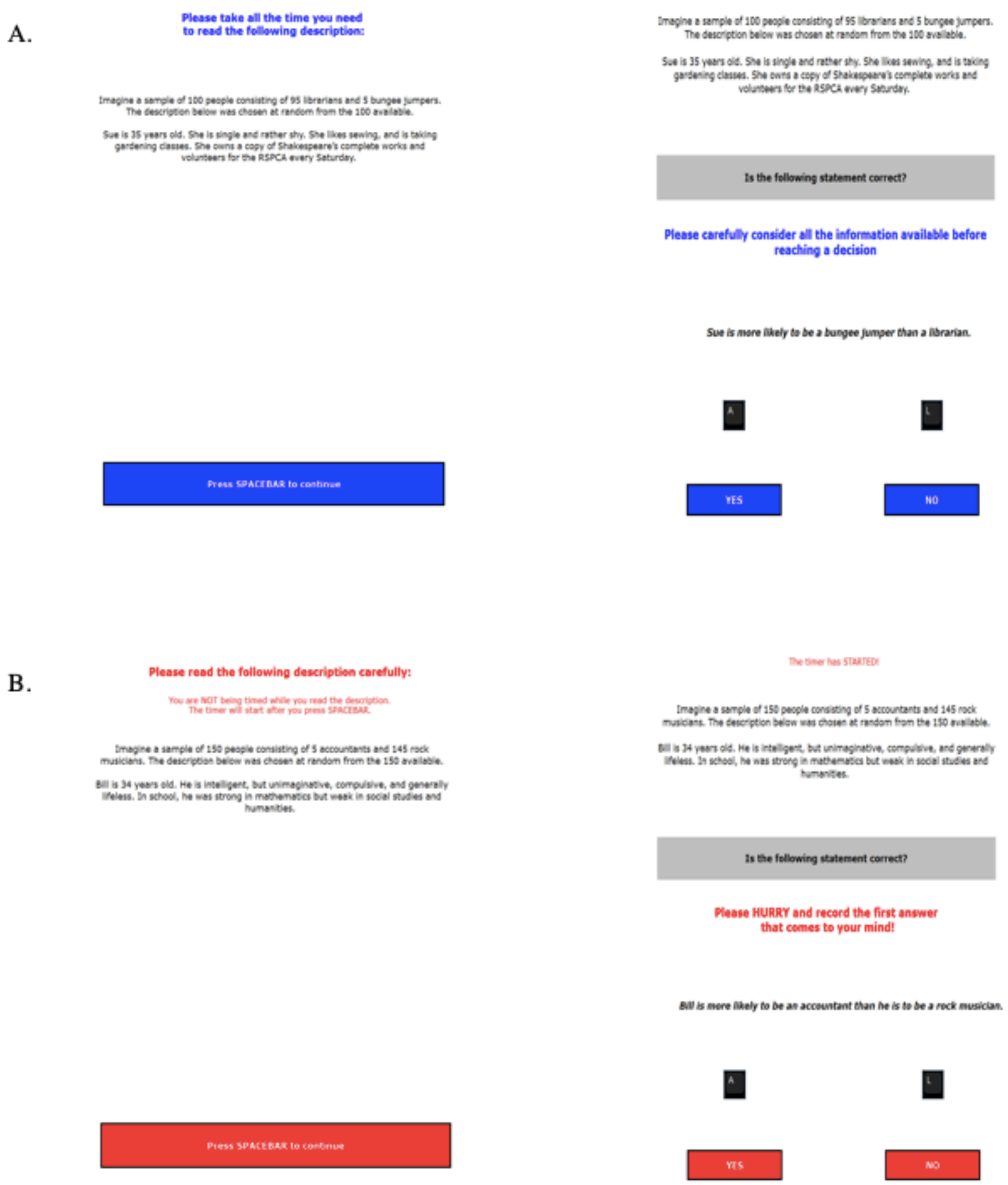


Figure 5. 1. Screenshots of the trials in both the unlimited time (A) and time pressured conditions (B). *Note*. A represents a U&I trial, while B represents a R/I trial

Results

Outliers. Following the procedures outlined by Meade and Craig (2012) I screened the sample for careless responders. An overall "flag score" was computed to identify careless responders who were identified as outliers on four different criteria. First, the total duration latency was converted in to a z-score. Outliers were identified as any cases with *z*-scores outside the (-3,3) range. Two participants were flagged as outliers in the no time pressure condition, and none in the time pressure condition. Second, average trial durations were computed for the time pressure and no time pressure conditions, these were then analysed for outliers. There was one participant flagged in the no time pressure condition, and one in the time pressure condition. Third, the data was screened for multivariate outliers in the judgement latencies. Mahalanobis scores showed that ten participants were flagged in the no time pressure condition for taking too long to make a judgement, while only one was flagged in the time pressure condition. Fourth, the data was screened for multivariate outliers in the reading times of the base rate and descriptive information. Mahalanobis scores showed that one participant was flagged in each of the time pressure conditions for taking too long to read the descriptive information. Next, a total flag score was computed for each participant. Two individuals who scored more than two flags were eliminated from the sample, leaving the sample size at $N = 77$. Finally, the sample was also screened for unbiased reasoners (i.e., participants who always chose the logical answer). One participant responded 100% logically to all 16 trials, thus was removed from the analysis. The final sample size, excluding all outliers, was $N = 76$ participants.

Manipulation checks. To check whether the manipulation of time pressure worked in this study I compared total durations across the two time conditions, and also compared confidence ratings and confidence latencies across these time conditions. Due to the design of the study, and every participant doing both the time pressured trials as well as the non-time pressured trials, I was expecting to see no difference in total time taken to complete the study regardless of the order they completed the conditions in. Indeed, this was the case. Participants who were put in the non-time pressure condition first completed the study in a similar time to those put into the time pressure condition first, $M_{NTPfirst}$ = 533.26s, *SD* = 147.39, 95% CI [484.12, 582.41] vs. $M_{TPfirst}$ = 509.01s, *SD* = 129.24, 95% CI [467.12, 550.91], $t(74) = 0.76$, $p = .447$, Cohen's $d = 0.18$. Judgement latencies were compared across time pressure conditions and the difference was significant. Participants judgements were faster for the time pressure trials than the non-time pressure trials, M_{NTP} = 5.19s, *SD* = 2.76, 95% CI [4.56, 5.82] vs. M_{TP} = 3.14s, *SD* = 1.40, 95% CI [2.82, 3.46], $t(75) = -6.86$, $p < .001$, Cohen's $d = 0.94$.

Participants were asked to rate their confidence in their judgements after each trial. Their answers as well as their judgement latencies were recorded. The idea behind this being that participants when put under time pressure would feel less confident in their judgements and would hesitate for longer when making these judgements, compared to when completing the non-time pressure trials. Indeed, there was a significant difference in confidence scores between the time conditions; however, surprisingly there was no difference in the time that they took to make these judgements. Paired samples t-tests revealed that participants were more confident in their answers when not under time pressure compared to when under time pressure, $M_{NTP} = 7.41$, $SD = 1.17$, 95% CI [7.15, 7.68] vs. $M_{TP} = 7.13$, $SD = 1.37$, 95% CI [6.82, 7.45], $t(75) = 2.38$, $p = .020$, Cohen's $d = 0.22$; however, there was no difference in latencies when making these judgements, M_{NTP} = 2.08s, *SD* = 0.80, 95% CI [1.90, 2.26] vs. M_{TP} = 1.99s, *SD* = 0.79, 95% CI [1.81, 2.17], $t(75) = 0.95$, $p = .344$, Cohen's $d = 0.11$.

Conflict Sensitivity. To examine patterns of heuristic responding, a heuristic score was computed for each statement type by calculating the proportion of time participants either accepted statements that were congruent with heuristic considerations (R&L and R/I statements; see Table 2. 1, p. 74) or rejected statements that were incongruent with such considerations (U/L and U&I statements). In this context, U/L means the participant answered in accordance with the base rate information (ignoring the personality description), thus making a logical judgement. R/I judgements are made in line with the representative personality description (ignoring the base rate information), thus making an illogical judgement. R&L and U&I statements contain no conflict between the base rates and the personality stereotypes. The heuristic score ranged from 0 (never followed the heuristic consideration) to 100% (always followed the heuristic consideration). Heuristic scores were compared across statement conflict to assess whether the presence or absence of conflict in the statements affected the level of heuristic considerations used by participants' when making base rate judgements. Heuristic scores were also compared across time conditions to assess whether the presence or absence of time pressure affected the level of heuristic responding. Finally, the order in which the participants completed the statements was included as a repeated measures factor (i.e., time pressure first, or unlimited time first). Although the task was separated into two distinct parts (time pressure or unlimited time), there might still be an order effect apparent, hence the inclusion of order as a repeated measures factor. By starting the task off under severe time pressure, one might be unknowingly setting the scene for the rest of the trials despite instructions to slow down and think hard about the answers. Likewise, by beginning the task with no time pressure, one might be lessening the urgency of the time pressure condition that follows. Thus, a 2 (conflict: absent vs present) x 2 (time pressure: unlimited vs limited) x 2 (order: limited-unlimited vs unlimited-limited) mixed ANOVA was conducted with repeated measures on the first two factors and independent measures on the last factor.

There was a significant main effect for conflict. Results showed that statements for which heuristic considerations conflicted with logical ones led to lower rates of heuristic responding than non-conflict statements, $M_{noconflict}$ = 92.78%, SD = 11.95, 95% CI [90.04, 95.53], $M_{conflict}$ = 85.39%, SD = 17.84, 95% CI [81.29, 89.49], $F(1, 74) = 13.44$, $p < .001$, η_p^2

= .15. Surprisingly, there was no effect of time pressure. Heuristic scores did not differ between those trials completed under unlimited time and those completed under time pressure, $M_{unlimited}$ = 88.80%, *SD* = 15.62, 95% CI [85.21, 92.39], $M_{limited}$ = 89.38%, *SD* = 14.25, 95% CI [86.15, 92.60], $F(1, 74) = 0.09$, $p = .761$, $\eta_p^2 = .001$. There was no significant interaction between conflict and time pressure condition, $F(1, 74) = 0.48$, $p = .491$, $\eta_p^2 = .01$ (see Figure 5. 2 for descriptive statistics). Furthermore, the order in which participants completed the trials (i.e., whether they performed the time pressure trials first, or the unlimited time trials first) made no difference to their heuristic scores, $M_{NTPfirst}$ = 90.03%, *SD* = 9.37, 95% CI [85.95, 94.12], $M_{TPfirst}$ = 88.14%, *SD* = 14.82, 95% CI [84.16, 92.12], $F(1, 74) = 0.44$, $p = .510$, $\eta_p^2 = .01$. Finally, there were no other significant interactions found in the data: conflict x order ($p = .883$), time pressure x order ($p = .100$), and conflict x time pressure x order ($p = .629$).

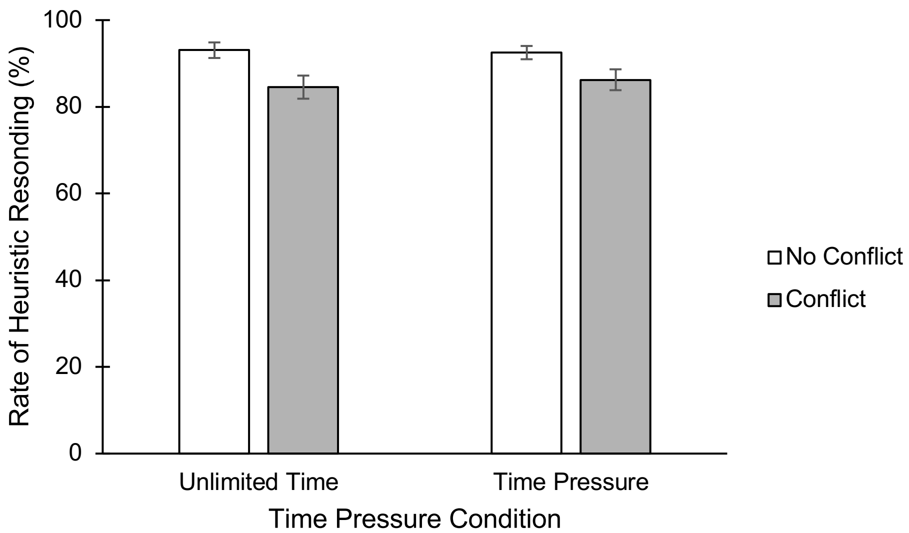


Figure 5. 2. Rate of heuristic responding as a function of the conflict between heuristic and logical considerations and the time pressure condition. *Note*. Error bars represent standard errors

There was also evidence for conflict sensitivity in the self-reported confidence scores. Following every judgement they made, participants had to rate how confident they were that their answer was correct. Their confidence ratings as well as their confidence rating latencies were recorded. Paired samples t-tests confirmed that confidence ratings were higher when conflict was absent from the statements, $M_{no\ conflict} = 7.49$, $SD = 1.14$, 95% CI [7.23, 7.75] vs. $M_{conflict} = 7.06$, $SD = 1.40$, 95% CI [6.74, 7.38], $t(75) = 3.64$, $p < .001$, Cohen's $d = 0.34$. There was no difference in confidence rating latencies between conflict and no-conflict statements, $M_{no\ conflict} = 2.01$s, $SD = 0.73$, 95% CI [1.85, 2.18] vs. $M_{conflict} = 2.05$s, $SD = 0.76$, 95% CI [1.88, 2.23], $t(75) = -0.54$, $p = .594$, Cohen's $d = 0.05$.

Judgement Latencies. Response latencies were anlaysed to assess which responses showed sensitivity to conflict between heuristic and logical considerations. Assuming that people who always respond with a heuristic judgement under conflict use a different strategy overall, I first screened and excluded these participants, so I could compare latencies for heuristic and logical answers under conflict ($n = 36$, 45%). In line with previous research (Pennycook, Fugelsang, & Koehler, 2012; De Neys & Glumicic, 2008) three average latency scores were computed for each participant: the logico-heuristic answers on non-conflict trials (i.e., only the latencies for accepting R&L or rejecting U&I were selected), the heuristic answers on conflict trials (i.e., they either accepted R/I statements or rejected U/L statements), and finally the logical answers on conflict trials (i.e., they either rejected R/I statements or accepted U/L statements). Latencies were converted to $\log^{10}$ prior to analysis to normalize the distribution and subjected to a 3(judgment type: no-conflict vs. conflict-heuristic vs. conflict-logical) x 2(time condition: unlimited vs. limited) repeated measures ANOVA.

As expected, there was a significant main effect for time pressure condition with latencies being faster in the limited time condition than the unlimited condition, $M_{unlimited}$ =

3.70, SD = 0.57, 95% CI [3.59, 3.81] vs. $M_{limited}$ = 3.50, SD = 058, 95% CI [3.40, 3.59], $F(1,15)$ = 13.42, p = .002, η_p^2= .47; see Table 5. 1 for descriptive statistics in seconds.

Table 5. 1

Means, Standard Deviations and 95% Confidence Intervals of judgement latencies as a function of time pressure condition. Note Means and Standard Deviations reported in seconds

Time condition	*M*	*SD*	95% Confidence Interval	
			Lower Bound	Upper Bound
Unlimited	6.21	3.55	4.45	7.98
Time Pressure	3.60	1.56	2.62	4.59

There was no significant difference in the latencies between types of responses, $M_{no\text{-}conflict}$ = 3.59, SD = 0.18, 95% CI [3.51, 3.67] vs. $M_{conflict\text{-}heuristic}$ = 3.61, SD = 0.34, 95% CI [3.49, 3.74] vs. $M_{conflict\text{-}logical}$ = 3.59, SD = 0.86, 95% CI [3.48, 3.69], $F(2,30)$ = 0.15, p = .860, η_p^2 = .01; see Table 5. 2 for descriptive statistics in seconds. Finally, there was no interaction between time pressure and type of response, $F(2,30)$ = 1.20, p = .154, η_p^2= .12, see Figure 5. 3 for descriptive statistics.

Table 5. 2

Means, Standard Deviations and 95% Confidence Intervals of judgement latencies as a function of response type. Note, latencies are reported in seconds

Type of Response	*M*	*SD*	95% Confidence Interval	
			Lower Bound	Upper Bound
No conflict	4.35	2.40	3.39	5.30
Conflict-Heuristic	5.51	3.04	3.14	7.87
Conflict-Logic	4.87	3.89	3.48	6.26

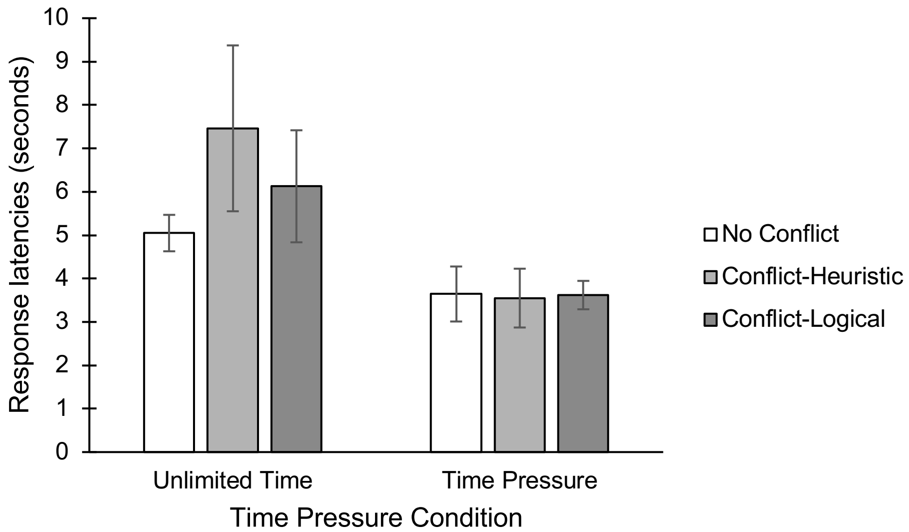


Figure 5. 3. Response latencies as a function of time pressure condition and type of response. Note. Error bars represent standard errors

Cognitive Reflection and Logical Sensitivity. To further understand the origin of the conflict between heuristic and logical assessments, individual differences in cognitive reflection (i.e., the ability to override a heuristic response and replace it with a logical one) were evaluated to see whether they would predict logicality. If, as suspected, sensitivity to logic is fast and intuitive, but subsequently overridden by biases when given time to deliberate; then people who present low CRT scores (i.e., they do not override the prepotent heuristic response) should also be more logical in their base rate judgements. A logical sensitivity score was computed for each participant. More specifically, one score that showed what percentage of time they judged conflict statements logically (0% = "never logical", 100% = "always logical"). The sample was screened for perfect logicality and participants who always answered logically on the conflict trials were excluded from the analysis. A Pearson's correlation coefficient was computed to assess the relationship between logical sensitivity and cognitive reflection. No significant relationship was found between these two variables, $r = 0.22$, $p = .053$. A scatterplot summarises the results (Figure 5. 4). Increases in logical sensitivity did not correlate with increases in cognitive reflection scores. Additionally, only one participant scored 3 on the CRT. If this participant is removed as an outlier, then the result remains not significant; $r = 0.22$, $p = .063$.

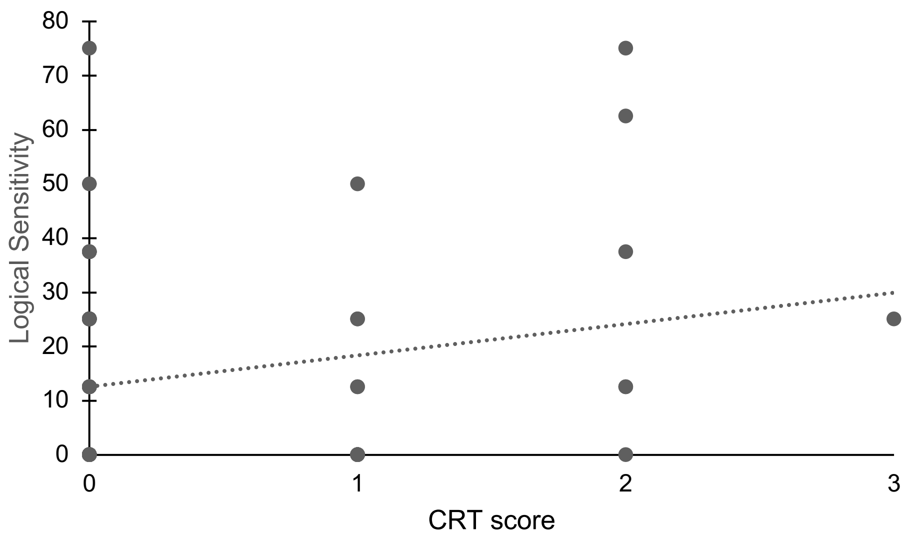


Figure 5. 4. Scatterplot of the relationship between CRT scores and logical sensitivity

Discussion

Support of Hypotheses and Interpretation of Results. The experiment reported here examined the impact of conflict and time pressure on the logicality of people's probability judgements. The results supported the hypothesis that people easily detect conflict between intuitive and deliberative considerations on base rate tasks, but there was no support found for logical intuitions. This was evident in the rates of heuristic responding, such that participants answered more in line with heuristic considerations when conflict was absent from the statements, but they were more logical in their judgements when conflict was present. This suggests that the presence of conflict in the base rate statements produced higher rates of logicality in the judgements, which is supported by findings in Villejoubert (2009) and Franssens and De Neys (2009). However, it is important to note that there is unfortunately no way to tell for certain why conflict spikes levels of logical responses in this study due to the methodology of the task. It could simply be that conflict causes a situation of uncertainty for the participants, which in turn causes the participants to make more guesses. The task forces a response between only two options; thus, a logical response could be the result of a) a logical consideration, b) the rejection of the heuristic response, or c) a guess. There was; however, evidence found for conflict sensitivity in the participants self-reported confidence ratings. They were more confident in their answers being correct when conflict was absent from the statements, than when conflict was present. Unfortunately, there was no support found for conflict sensitivity in the judgement latency data. Contradictory to Experiment 1, the three different types of responses given by participants (e.g., no-conflict, conflict-heuristic and conflict-logical) were all made in similar judgement times to one another. These findings do not support the idea of conflict detection, as the no-conflict statements were answered just as quickly as those containing conflict, suggesting that perhaps the participants did not detect the difference between the two, or perhaps they did but were too "miserly" to spend any extra time or cognitive energy on these trials.

Perhaps the biggest surprise in this study was the fact that time pressure had no impact on the participant's logicality. Participants judged eight statements under time pressure, and an additional eight statements without any time pressure, and there were no differences found in rates of heuristic responding between these two conditions. Although the manipulation was successful, and participants judged statements significantly faster when under time pressure, there was no evidence that this translated into the types of reasoning they used when making these judgements. As seen in Villejoubert (2009) and Kahneman (2011), time pressure was employed to encourage System 1 thinking resulting in higher intuitive responses, while unlimited time would allow System 2 thinking resulting in more deliberative responses (e.g., as was the case in Experiment 1). However, in this study there was no evidence for different thinking styles across the two time conditions. Participants were equally heuristic in their responses ($M_{unlimited}$ = 88.80% vs. $M_{limited}$ = 89.38%) across both the time conditions. It is always possible that one or the other specific task or method feature is driving the effects in each specific task. For example, the role that conflict plays in logicality. Conflict between heuristic and logical considerations on ratio bias tasks was found to result in fewer correct responses (Bonner & Newell, 2010), while conflict on base rate neglect tasks, and conjunction fallacy tasks resulted in more correct responses (Franssens & De Neys, 2009; Villejoubert, 2009). However, as convergent evidence across different tasks and methods increases, the possibility that the effect is being driven by a specific task or method becomes more unlikely (De Neys, 2013). The results from this experiment lend no support for logical intuitions. Time pressure had no impact on the logicality of the responses, in fact, the rate of heuristic responding remained higher than any found across this body of research. Either the time pressure manipulation failed to induce System 1 thought, which might be due to the nature of the task (for example, base rate tasks might require less time to answer than conjunction fallacy tasks, thus even a time limit of 7.5 seconds was too long for System 1 reasoning. However, this is unlikely as researchers have presented response times much slower than these in their results concerning base rate tasks, which will be discussed in more detail in the following section). Or, the participants were too miserly with their cognitive resources to deliberate their answers, hence the surprisingly fast responses across both of the time conditions. The same way that there is no support found for logical intuitions, there is no support for heuristic reflections, due to the failure of the time pressure manipulations to encourage System 1, or System 2 reasoning. There is no difference in responses between the two time conditions suggesting that the participants employed a similar reasoning style across both the conditions instead of switching between intuitive and deliberative.

Finally, there was no support for the hypothesis that people with little or no cognitive reflection also show more logicality in their judgements. In fact, if anything, the data seemed to show a trend moving in the opposite direction: an increase in cognitive reflection resulted in an increase in logicality. However, this finding was not significant revealing no relationship between cognitive reflection (i.e., the ability to inhibit an intuitive response and replace it with a deliberate one). It is also important to note how uneven the groups were with only one participant answering all three CRT questions correctly. Below is a breakdown of the responses, which shows how few participants were able to answer any of the difficult CRT questions correctly. The majority of participants never got any answers correct (n = 57, 75%), and only one participant managed to answer all three questions correctly (see Figure 5. 5).

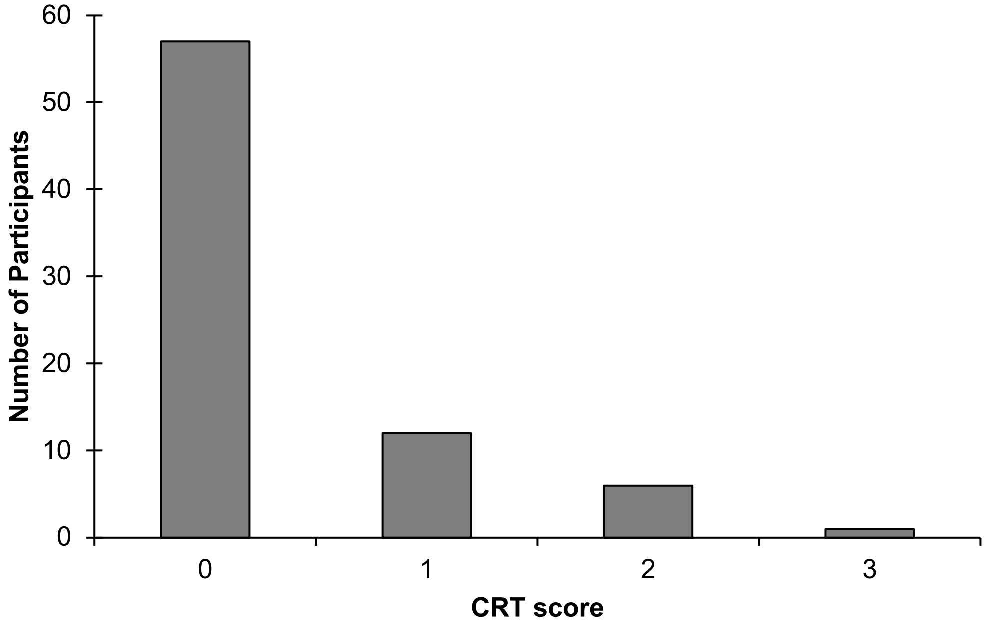


Figure 5. 5. Distribution of participant's CRT scores (N = 76)

Toplak, West and Stanovich (2011) demonstrated that the CRT was a powerful predictor of heuristics and biases tasks. In fact, they suggest that cognitive miserliness is the primary factor in predicting how well people perform on the task. However, Stupple, Gale and Richmond (2013) found that while cognitive miserliness was implicated in CRT performance, it was in fact working memory capacity that was the strongest predictor. There is a recent stream of thought that believes the CRT is actually a test of numeric ability (compared to general intelligence) as one requires numeric ability to solve the problems correctly (Sinayev & Peters, 2015; Welsh, Burns & Delfabbro, 2013). These researchers have found that the CRT positively correlates with numeracy and that it's predictive power is limited to biases with a numerical basis, such as framing and discount delay tasks. Base rate tasks also have a calculable correct response; however, Welsh et al. (2013) found no significant relationship between base rate tasks and the CRT. Indeed, numeric ability is a robust predictor of superior decision making and seems to be the key mechanism (compared to cognitive reflection) when solving the CRT problems (Sinayev & Peters, 2015). If cognitive reflection (i.e., the ability to override a prepotent intuitive response) is not actually being measured by the CRT, then it is not surprising that the hypothesis was not supported. Instead, what the results might be showing is a weak, non-significant relationship between numeracy and logicality. However, the relationship between numeracy and good judgement making would need to be tested further. Taken together, these findings highlight the fact that the psychological and psychometric properties of the CRT require further research.

Generalizabitily of Results. Problems with extreme base rates? De Neys and Glumicic (2008) used extreme base rates (e.g., 995/5) and found evidence of conflict detection. Using less extreme base rates (e.g., 95/5 or 9/1) in other studies (De Neys & Franssens, 2009; De Neys & Feremans, 2013) still resulted in successful conflict detection. Thus, extreme base rates do not appear to affect conflict detection (De Neys, 2014). However, Pennycook et al. (2012) performed an extensive set of experiments and found that conflict detection, as represented by latencies, is more pronounced when more extreme base rates are being used. Indeed, even the biased reasoners took longer to make a judgement when conflict was present in the trials. However, when they tested with the classic 70/30 base rate the effect disappeared. The researchers suggested that the more extreme base rates help to focus attention on the base rate information and thereby facilitate conflict detection. The current study used a mix of base rates; however, they all erred on the side of extreme and the difference in the ratio was always fairly large. In fact, the least extreme base rate used was 175/25 (e.g., 88:12), while the most extreme was 985/15 (e.g., 99:1).

The question remains: why did the participants not show conflict detection in their judgement latencies in this study? They made no-conflict, conflict-heuristic and conflict-logical judgements in a similar time to one another, even when coerced to answer quickly or leisurely, which is extremely unusual considering the fact that the differences in latencies were significant in all of the other experiments across this program of research (except for Experiment 2 which was underpowered). The answer to this might have something to do with the design of this study (e.g., repeated measures) or the task itself (e.g., base rates) that has made the participants answer the unlimited time condition very quickly. Compared to Experiment 1, where participants in the unlimited time condition took on average 19 seconds to make a judgement (while the limited time condition took 4.38 seconds); participants in the current study took 6.21 seconds (and 3.60 seconds in the limited time condition). Although this difference is significant which indicates they were indeed faster when under time pressure, they were not exactly slow when given unlimited time. In fact, if you deduct 1.65 seconds for the initial reading time of the 11-word statement, they judged their "slow and deliberative" answers in a mere 4.56 seconds. The task was randomized and counterbalanced and there were no order effects found in the data analysis, which leads me to speculate that the speed at which they completed the judgements had to do with the type of task being administered (e.g., base rates instead of conjunctions) and not the study's design (e.g., repeated measures). On closer examination of the latency data, it became apparent that perhaps the independent grouping condition did in fact alter the latencies, as latencies for the time pressured judgements appeared faster if participants were in the no-time-pressure-first group. However, a 2(time pressure: unlimited vs. limited) x 2(group: NTPfirst vs. TPfirst) one way ANOVA revealed a non-significant difference in latencies between the no time pressure and limited time pressure conditions as a function of grouping condition. Firstly, there was no significant difference in judgement latencies for the time pressured trials as a function of group, $M_{NTPfirst}$ = 2.83s, SD = 1.64, 95% CI [2.29, 3.38] vs. $M_{TPfirst}$ = 3.43s, SD = 1.07, 95% CI [3.08, 3.78], $F(1,74) = 3.53$, $p = .064$. Nor was there a difference in judgement latencies for no time pressure trials as a function of group, $M_{NTPfirst}$ = 5.59s, SD = 2.47, 95% CI [4.76, 6.41] vs. $M_{TPfirst}$ = 4.81s, SD = 3.00, 95% CI [3.84, 5.78], $F(1,74) = 1.51$, $p = .223$. See Figure 5. 6. Thus, the group (e.g., time pressure first, or unlimited time first) was not responsible for the fast response latencies as there were no differences found between the two.

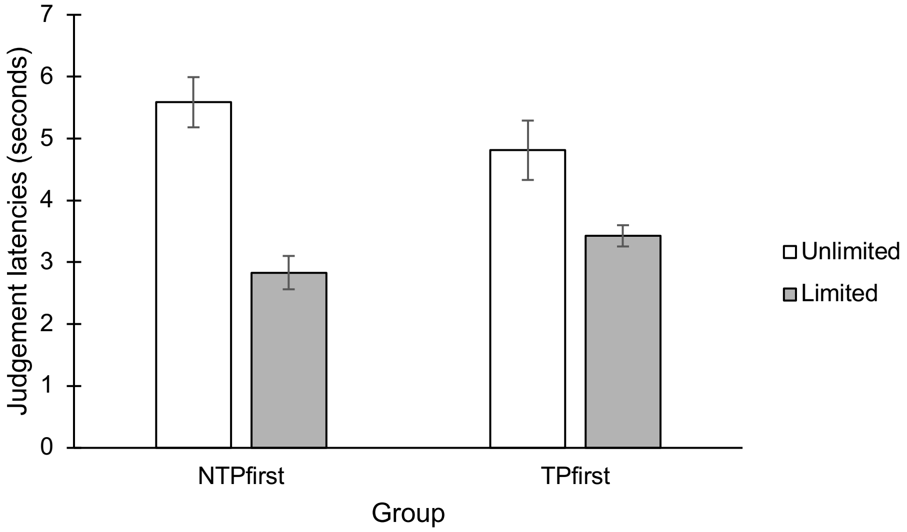


Figure 5. 6. Judgement latencies as a function of time pressure and group. *Note*. Error bars represent standard errors

De Neys and Glumicic (2008) conducted two experiments using base rate tasks. They reported their participant's response times as ranging from fastest at 14 seconds (when responding correctly to congruent trials) to slowest at 24 seconds (when responding correctly to incongruent trials). Pennycook, Trippas, Handley, and Thompson (2013) also conducted a series of experiments using base rate neglect tasks. In experiment one, they found a main effect for instruction prompts (e.g., belief vs. statistics) such that statistics instructions (M = 12.73s) led to faster judgement latencies than belief instructions (M = 14.05s). In experiment two, they found that participants responded faster to no conflict (M =3.69s) than to conflict (M = 3.80s) trials. In experiment three, they do not disclose latency data. Experiment two was under time pressure while experiment one was not. The judgement latencies from these two research articles appear 'normal' in the sense that unlimited time judgements took a long time to make (over 12 seconds), while time pressured judgements were significantly shorter in under 4 seconds. Which is indeed was found in Experiment 1, which compared conjunction judgements made under time pressure to those made under no time pressure (note, this study was an independent measures design). Participants took around 19 seconds to make no time pressure judgements, and they were much faster (4.38 seconds) when making time pressured judgements. Then why were participants in the current study so fast to make their unlimited time judgements ($M_{unlimited}$ = 5.20s vs. $M_{limited}$ = 3.13s)? The instructions were the same as Experiment 1, there were blue colour prompts as well as wording that requested participants to think hard about their answers in the unlimited time condition (e.g., "please carefully consider all the information available before making a decision"), while the red colour prompts and urgent wording in the time pressure condition (e.g., "please HURRY and record the first answer that comes to your mind"), as well as the risk of losing credits, ensured participants answered quickly in the limited time condition. A possible reason for this latency discrepancy might be a result of the briefing that participants were given before commencing with the task. Before they began, they were informed that they would be participating in both time conditions: 8 trials under time pressure and 8 under no time pressure, but the order of these would be randomized. They were also informed that they would lose credits if they took too long to make a judgement in the time pressure condition. Providing this information before the study might have somehow compelled the participants to make fast judgements in both conditions. Future studies might also include a short break between the time conditions, as a way to "reset" the participants of any learned behaviour from the first condition.

Final remarks and Future Directions. This study was a conundrum. On the one hand it showed a perfect example of how conflict effects logicality in probability judgements. Participants were sensitive to the conflict underlying intuitive and deliberative assessments, and the presence of conflict caused them to be more logical in their judgements. This supports the findings in Villejoubert (2009) and Franssens and De Neys (2009). While on the other hand it showed that time pressure made no difference to individuals' logicality. In Experiment 1, the presence of time pressure significantly increased individuals' logicality, and this was expected to be replicated in the current study; however, it was not. In the current study all participants answered with high levels of base rate neglect regardless of whether they were under time pressure or not. The evidence of conflict detection combined with high base rate neglect rates provides support for the parallel-competitive account. The two systems of thought run concurrently allowing participants to detect when their answers conflict with one another. When this happens, it allows the participants the opportunity to employ further deductive reasoning which might result in System 2 inhibiting System 1. However, this is not always the case, as the heuristic response can be very compelling, hence the high rates of fallacies committed.

A further strange finding was that participants in this study completed the task extremely quickly, independent of time pressure condition. In order to deduce whether this was due to the design of the study being within subjects, or due to the task itself, it would be beneficial to conduct a few replication studies in the future. The first could be to conduct the study as a between subjects design (i.e., participants are allocated to either the time pressure or the unlimited time condition, and not both) to see whether this makes a difference to the logicality scores. Additionally, the within subjects design could be replicated but this time should include less transparent instructions before the task commences. Instead of briefing the participants in full before the task begins, perhaps the instructions could unfold only at the beginning of each section of the task. I would also suggest making it explicitly clear that credits are only lost due to slow responses in the time pressured condition, and that the unlimited time condition is completely separate from this. I suspect that participants were probably quite anxious at the prospect of losing credits (e.g., money) and as a result responded quickly to all items in the task (i.e., both time pressure conditions). I would also recommend taking a short break, or introducing a filler task, between the two conditions as a way to eliminate passing any learned behaviour from one condition to the next. However, these suggestions are purely speculation will need to be tested with further research.

Chapter 6: General Discussion

The conjunction fallacy has historically been explained through dual process theories, such that the reliance on heuristics – specifically representativeness – produces a bias in people's judgements. In other words, System 1 processes, or intuition, leads people astray. However, if people were only able to inhibit the pervasive intuitive response and employ system 2 deliberation, they would arrive at a correct, logical answer (Tversky & Kahneman, 1983; Evans, 2008; Sloman, 1996; Stanovich & West, 2000). This is due to the fact that traditionally System 1 processes are viewed as rapid, autonomous processes and are assumed to yield default biased responses unless intervened on by distinctive higher order reasoning processes (System 2). What defines the difference between the systems of thought, is that System 2 processing supports hypothetical thinking and load heavily on working memory (Evans & Stanovich, 2013). At the heart of the dual process model of the Linda problem lies a "corrective" view on sound reasoning, such that accurate or logical responding is assumed to require the correction of an intuitive System 1 response by slower, more demanding System 2 processing (Kahneman, 2011; Kahneman & Frederick, 2005). This view leads to a somewhat unattractive characterization of System 1 as a source of error that requires supervision from the deliberate System 2[2].

While the existence of two systems of thought (i.e., intuition vs. deliberation) is plausible, the emergence of recent work on conflict detection and logical intuitions calls into question the role played by intuition in conjunctive probability judgment. It is apparent that

[2] It should be clarified that dual process models do not simply claim that intuitions are always incorrect. It is not disputed that intuitive responses can be appropriate and helpful in some cases (Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996). It is also not claimed that deliberation will necessarily lead to a correct answer (Evans & Stanovich, 2013; Kahneman, 2011). The point is that intuitive responses have the potential to be incorrect, thus require monitoring and sometimes correction. It is this corrective view that leads to a rather negative view of System 1.

people's intuition does not always lead them astray, indeed in certain instances it leads them to the correct logical response (De Neys, 2012; Villejoubert, 2009; Pennycook, Trippas, Handley, & Thompson, 2013). People are also readily able to detect the conflict between intuitive and deliberate processes, even when they are biased. In other words, they can detect when the heuristic System 1 response is at odds with the logical System 2 response, even if they are not able to suppress it (Pennycook, Fugelsang, & Koehler, 2012; Franssens & De Neys, 2009, Villejoubert, 2009). These results pave the way for an alternative explanation, where logical judgements result from the absence of overriding the logical intuitive assessment. This view is closest to the logical intuitionist account as it assumes logical assessments arise from intuitive processing; however, it also proposes that heuristic judgements result from overriding logical intuitive assessments through engaging in reflective thinking. Consequently, heuristic judgements (unlike heuristic assessments) require the involvement of reflective thinking. This idea was tested across a series of five experimental studies using an adapted methodology from Villejoubert (2009).

Overview of Hypotheses and Results

This thesis had two objectives. Using the dual-process theoretical framework, the first explored the impact of individual differences on the logicality of peoples' probability judgements. Experiment 1 attempted to clarify whether people have access to logical intuitions (e.g., the logical answer originates from intuitive processing), but these are subsequently overridden by heuristic deliberations when they think hard about their answer. Additionally, it was investigated whether individuals who engaged more naturally in effortful thinking (i.e., achieved a high score on the rational ability, and a low score on the experiential ability) were also more prone to produce illogical probability judgements (i.e., a high heuristic score) if they had time to deliberate because their deliberation was biased by non-logical heuristics (e.g., prior beliefs). Experiment 2 tested whether in situations of conflict between representativeness and logicality (e.g., R/I and U/L statements), individuals who provided logical answers would show different eye-movement patterns to individuals who provided heuristic answers. This tested for evidence of conflict sensitivity and explored whether certain strategies in obtaining information from the task (i.e., as interpreted from the eye movement patterns), resulted in higher or lower logicality in the participants answers.

The second objective attempted to test original and theoretically driven ways to improve individuals' probability judgements and explored whether these were transferable across heuristics and biases tasks. Experiments 3 and 4 tested the hypothesis whether decreasing individuals' capacity for effortful thinking also had a positive impact on probability judgements because it impeded the inhibition of the rapid, logical response by conscious effortful deliberation. While Experiment 5 tested the transferability of the findings from Experiment 1 to a different heuristics and biases task (e.g., base rate neglect). It attempted to replicate findings from Experiment 1, by testing the assumption that both the presence of conflict and severe time pressure would increase logicality in base rate judgements in a within-subject design.

Experiment 1. The first experiment examined the impact of time pressure on the sensitivity to the conflict between heuristic and logical assessments in social judgements under uncertainty. It also explored whether a natural propensity to think rationally, or experientially made a difference to people's logicality when they made conjunction judgements. The hypothesis proposed that people could readily detect conflict between the two systems of thought. This was supported by results that showed conflict-statements were associated with lower rates of heuristic judgements and longer judgement latencies. Importantly, time pressure had an impact on logical judgements, such that people made more logical judgements when conflict was present in statements, and this effect increased under time pressure. These findings do not provide support for the default interventionist account. The fact that people exhibited sensitivity to conflict in their response latencies, plus this translated to higher logicality in their judgements, instead suggests support for the parallel competitive account of reasoning. In other words, the two systems of thought run concurrently thus being able to alert people to conflicting answers between them. This conflict detection results in system 2 overriding the intuitive response in some cases; hence the higher rate of logical/correct answers when conflict was present. However, the results might also point towards the more recent logical intuitions account. Participants made more logical judgements when conflict was present and this effect was increased when put under severe time pressure, suggesting that fast System 1 thinking might in fact reflect logical considerations. When given unlimited time to think about their answers (as exhibited by their response latencies), participants seemed to respond more in line with heuristic considerations, suggesting that slower deliberation does not result in more logical judgements, and can be influenced by non-logical biases, such as prior beliefs. However, this hypothesis was debunked during the analysis of the REI data. There was still support for the logical intuitions account as participants under severe time pressure responded significantly more logically compared to those in the unlimited time condition, but there was no evidence that pointed to the heuristic bias originating in System 2 reasoning. Instead, the marginally significant evidence suggested that individuals who rated themselves as medium on both rational and experiential scales actually outperformed those who rated themselves as high on both scales. This finding does not support the heuristic reflections hypothesis, which states that people who exhibit a natural propensity to deliberate (e.g., a high rational score) will also display more heuristic errors, because biases such as prior beliefs corrupt people's deliberation.

Experiment 2. This experiment examined the hypothesis whether in situations of conflict between representativeness and logicality, individuals who provided logical answers would show different eye-movement patterns to individuals who provided heuristic answers. It also explored the relationship between working memory capacity and logicality. To be more specific, participants who exhibited lower working memory capacity were expected to also exhibit higher logicality in their probability judgements because they would rely more heavily on their intuitive processing (i.e., produce logical intuitions). The behavioural results for this experiment did not yield much in the line of proof for the hypotheses, probably due to the fact that the study was severely underpowered. However, the eye-tracking data showed a natural ability to detect conflict, thus supporting the parallel-competitive account. Participants refixated more often (i.e., they looked back more often) when there was conflict present compared to when conflict was absent. There was also a significant interaction between conflict and AoIs for both number of refixations and dwell times. Participants searched for answers in the description when conflict was absent from statements; however, when conflict was present they searched for answers in the statements. This is very interesting as it shows that on some level, probably outside of their awareness as eye movements are unconscious, people realise that the answer to the no-conflict statements can be found in the personality descriptions; however, the answer to the conflict statements lies beyond the personality vignette and instead concerns the conjunction rule, hence they search the statements for their answer. There was also evidence found that the longer people refixated on the descriptions, the less logical their judgement were. This was not surprising as the description was designed to introduce a highly convincing stereotype. The more time a person spent contemplating the stereotype, the more likely they would be to answer in line with the representativeness heuristic. The results for experiment 2 provide no evidence of logical intuitions nor of heuristic reflections. There was a pattern found in the correlation between logicality and task duration that showed fast participants made more logical judgement; however, this relationship was not significant (p = 079). Perhaps with more power and a larger sample size this would reach significance, but further testing is needed. Finally, the hypothesis that a high working memory capacity would result in lower rates of the conjunction fallacy was not supported. It was found that having a strong, or weak, working memory capacity made no difference to the logicality of probability judgements.

Experiments 3 and 4. The third and fourth experiments explored ways of improving individuals' probability judgements. In line with the logical intuitions account, they examined whether depleting and overburdening cognitive capacity (i.e., discouraging effortful, deliberative thinking) had a positive impact on conjunction probability judgements because it encouraged fast, intuitive and logical assessments. Additionally, the experiments investigated the relationship between impulsivity and logicality, as well as cognitive reflection and logicality on social judgements. In accordance with the heuristic reflections account, it was hypothesised that low impulsivity, but high cognitive reflection would result in more conjunction fallacies because people displaying these traits would also be more inclined to deliberate, which is when the heuristic biases reasoning. The hypothesis was supported by the data that showed both cognitively depleted and overburdened participants were able to detect the conflict underlying the statements. This finding was confirmed by both the judgement latency as well as the heuristic scores data. Participants were faster on the no-conflict trials compared to the conflict trials. They were also responded heuristically (i.e., in line with heuristic considerations) in the absence of conflict. In line with the logical intuitionist account, both the depleted and overburdened participants responded with higher logicality in their probability judgements; suggesting that logical assessments might originate from intuitive processes, while the slower, more deliberative processes might be susceptible to heuristic biases. However, the methodology used in this experiment has some flaws and requires further testing to be conducted in this regard. Finally, there was no relationship found between impulsivity and logical sensitivity, neither was there a relationship between cognitive reflection and logical sensitivity.

Experiment 5. The final experiment tested whether the findings from the first experiment were robust across a within-subjects design as well as a different cognitive fallacy, namely base rate neglect. Based on the findings from Experiment 1, the hypothesis predicted that both conflict and time pressure would increase logicality in base rate judgements. Additionally, the relationship between cognitive reflection and logical sensitivity was explored. It was expected that participants who scored high on cognitive reflection also committed more conjunction fallacies, as their ability to inhibit intuition and employ deliberation would result in more biased answers (e.g., the heuristic reflections account). The data supported the hypothesis that participants were readily able to detect conflict between heuristic and logical considerations. They displayed higher logicality in their judgements for statements containing conflict, they were also less confident in their answers when conflict was present. However, there was no evidence found for conflict sensitivity when analysing the judgement latencies for the three different response types. This was an unusual result as latencies were significantly faster when put under severe time pressure (i.e., the time pressure manipulation appeared to work, but made no difference to the participant's answers in this experiment). Finally, there was no support for the hypothesis that a high cognitive reflection score would lead to more conjunction fallacies. In fact, an opposite pattern to this was emerging in the data, although it was not significant. Thus, having high or low cognitive reflection (e.g., the ability to inhibit intuitive responses), makes no difference to the considerations you rely on when making social judgements.

Contribution to Knowledge

Evidence for Logical Intuitions and Heuristic Reflections. The findings in Experiment 1 did lend support for the idea of logical intuitions and heuristic reflections; however, these need to be interpreted with caution due to certain methodological flaws found in this study. Participants displayed significantly more logic in their judgements when put under severe time pressure compared to when given unlimited time to reflect on their answers. When allowed enough time to deliberate, participants made more conjunction errors than when they were forced to respond with their intuitive answer. These findings suggest that the logical answer might come first, but when allowed to deliberate the heuristic bias then corrupts reasoning. However, the CCPJ task needs to be redesigned and modified to include multiple choices, which will eliminate the possibility of guessing or being forced to choose one of only two options.

Experiment 2 showed that eye movement patterns could predict logicality on conjunction judgements, and that people were sensitive to the conflict underlying intuitive and logical considerations. The longer participants spent looking at the descriptions (as measured by their refixation dwell times), the more conjunction errors they committed. While the amount of time spent looking at the statements made no difference to their judgements. If automatic and superficial levels of processing have shorter fixation latencies, while deeper processing and a deliberate consideration of information is related with longer fixations (Glöckner & Herbold, 2011; Velichkovsky, Rothert, Kopf, Dornhofer & Joos, 2002), then we can presume that longer refixation dwell times on the description also meant a deeper processing of that information. Hence, a deep processing of the information on the description also led to more biased judgements. This finding is both common sense and profound. The Linda problem was designed to illicit the conjunction fallacy. It exacerbates the fallacy by priming people with information that leads them to choose the conjunction as the more probable option. The point is that there is always a stereotypical response and a surprising response, and the representativeness heuristic is used to manipulate these responses in the description vignette. Thus, it is not that surprising that when people pay more attention to the description, they are also more influenced by the heuristic response. According to the traditional parallel-competitive model of dual process theories, people can detect conflict but are not always able to override the potent intuitive response. However, sometimes the presence of conflict can cause them to employ deliberation which can lead to a correct and logical answer. This was not the case according to the data in this experiment. In this case slow, deliberative reasoning (i.e., the lengthy refixations that suggested deliberate considerations) resulted in more heuristic responses (i.e., more conjunction fallacies). This evidence lends support for the heuristic reflections account that proposes the bias corrupts our reasoning when we reflect or deliberate. However, it could also support the lax monitoring and cognitive miser accounts as the participants might have been simply using their intuition to solve these problems, and not employing deliberation at all. Although, this could be disputed by the long reaction times recorded, which were arguably too long for supposedly fast, automatic responses.

Experiment 3 used the letter-*e* task to deplete participants' ego, or cognitive resources. The high ego depletion manipulation resulted in significantly fewer conjunction fallacies compared to the control condition. People who are cognitively depleted theoretically do not have access to the resources needed for demanding deliberative processing, thus they would tend to rely more heavily on their intuitive processing. This indicates that a higher reliance on intuition also resulted in fewer conjunction errors, which provides evidence for logical intuitions. Participants in the control condition would have had full access to the cognitive resources needed to conduct deliberative reasoning. However, despite this, they ended up committing more conjunction fallacies than those in the high depletion condition. This does not support the traditional default interventionist account, or parallel-competitive accounts. Instead, this result provides support for the heuristic reflections account, as it demonstrates that people who have the means and ability to think hard about their answers, are influenced by biased reasoning such as prior beliefs, and make more errors. It also lends support for the idea of logical intuitions, as the more depleted the participants were, the more logical their answers were. This result was replicated in experiment 4. Cognitive resources were limited in this study by overburdening the working memory. This yielded similar results, showing that the participants who's working memory was overburdened the most also committed significantly fewer conjunction fallacies compared to the control condition. These findings again support the logical intuitions and the heuristic reflections account, as participants who were forced to answer intuitively showed more logicality in their judgements, while those who deliberated their answers showed higher levels of error. This suggests that the logical answer comes first and can be intuitive, while the biased heuristic answer comes second when people deliberate. The results form Experiments 3 and 4 need to be interpreted with caution for two reasons. First, the CCPJ task does not allow the researcher to assess whether a logical judgement resulted from logical considerations, or the rejection of the alternative answer, or from simply guessing. The second, is that there was no new control group data collected for these experiments. They both used the unlimited time condition from Experiment 1 as their control group. This did not allow the control group data from Experiment 1 to be replicated, which would have greatly strengthened the evidence here.

Finally, experiment 5 did not yield any support for logical intuitions or heuristic reflections. The participants answered the base rate problems extremely quickly in general compared to the conjunction judgement problems, and although there was a significant difference in time between the trials answered under time pressure and those under no time pressure, this had no effect on the logicality of the base rate judgements. These results suggest that the participants might have been answering with one particular strategy in mind. For example, in line with the parallel-competitive account, they might have noticed the conflict between the two opposing responses as represented by the significant differences found in their heuristic scores and confidence scores; however, found the intuitive response too persuasive. Thus, the fault would be an inhibition one, which ties in with the cognitive miser account. This finding was indeed a surprise as the time pressure manipulation had worked so effectively to increase intuitive responses in Experiment 1, as well as other published studies (Villejoubert, 2009; Pennycook & Thompson, 2012; Bago & De Neys, 2017).

Evidence for conflict sensitivity. Across the program of research there was strong evidence to show that people are sensitive to the conflict underlying heuristic and logical considerations. In fact, conflict sensitivity was evident in all five of the experiments as represented by significant differences in heuristic scores, patterns in eye movements, or confidence judgements. Across the five experiments, conflict caused higher rates in logical reasoning, although this pattern was always visible these differences did not always reach statistical significance. In experiments 1, 4 and 5; however, a significant difference was found between conflict and heuristic scores, such that the presence of conflict caused lower rates of heuristic responses. In other words, the presence of conflict in the statements caused participants to make fewer probability errors, But it is important to note that across all five experiments the rates of heuristic responses remained very high (see Table 6. 1 for an overview). Experiments 2 and 3 did not show significant evidence for conflict sensitivity in the heuristic scores, however they did show evidence of conflict sensitivity in the eye movements in Experiment 2, such that participants refixated more on the description when conflict was absent from the statements, but refixated more on the statements when conflict was present. To be more precise, when conflict was present in the statements, the participants employed different eye movement patterns to when conflict was absent from the statements. They looked back more often and for longer at the statement when conflict was present, but they looked back and for longer at the description when conflict was absent. These results are fascinating as they show that people are very good at detecting conflict. In fact, conflict sensitivity might even occur involuntarily, as eye movements are uncensored, extremely fast and automatic. These findings also suggest that on some level people might have an idea or feeling as to the correct and incorrect answer because they appear to have an inkling of where to search for the correct answers. When there is no conflict, both systems of thought point to the same answer, hence the correct answer can be found in the description. However, when conflict is present the law of conjunction probability needs to be followed, thus looking at the statement is more likely to provide correct answers as it removes the compelling stereotype. Experiment 3 provided evidence of conflict detection in the self-reported confidence ratings provided by the participants after each trial they completed. Participants felt more confident in their answers when conflict was absent from the statements. Thus, again they were able to sense when their competing systems of thought were in conflict with one another and this translated into the amount of confidence they felt in their answers.

The fact that participants showed a sensitivity to the conflict underlying logical and heuristic considerations in all five of the experiments, supports the idea that conflict is easily and readily detected by people even if this is on an implicit and automatic level (e.g., as shown in uncensored eye movements in Experiment 2). These findings do not support the default interventionist account, which states that conflict detection is available only when a person engages in deeper System 2 processing.

Table 6. 1

Means, Standard Deviations, p-values and Effect Sizes of heuristic scores as a function of conflict across the program of research. Note, Means and Standard Deviations are expressed as percentages

Experiment	*M* (*SD*)		*p*	Effect size
	No conflict	Conflict		
1	86.67 (15.03)	78.76 (21.38)	< .001	$\eta_p^2 = .12$
2	86.59 (16.39)	83.23 (22.12)	.343	$d = .17$
3	85.24 (15.93)	83.31 (18.92)	.323	$\eta_p^2 = .01$
4	86.68 (14.50)	80.99 (18.87)	.001	$\eta_p^2 = .08$
5	92.78 (11.95)	85.39 (17.84)	<.001	$\eta_p^2 = .15$

Emerging Themes

Two interesting findings to come from this research program were that people want to think hard about their answers, and they are readily able to detect conflict between heuristic and logical considerations. The data across the program of research showed that people are effortless deliberators and not the “cognitive misers” who fail to reason in line with the tenets of logic because they either lack the cognitive ability or the motivation to do so (Kahneman & Tversky, 1973). To elaborate, the people who participated across these five experiments all showed a natural propensity to deliberate. They seemed to naturally, without coercion, engage in System 2 reasoning and think hard about their answers. In fact, it was increasingly difficult to get people to respond intuitively to the CCPJ task. This was particularly evident when designing and implementing the first experiment, which manipulated time pressure. All the participants, even those who were put under severe time pressure, appeared to want to think hard about their answers. It took many pilot versions of the experiment, including strict instructions to respond quickly, a countdown clock showing time running out, a stern vocal instruction to “answer now” to eventually arrive at the current design that uses the loss of money as an incentive to answer quickly. The lengths that the researcher had to go to get participants to answer as quickly as possible suggests that people tend to want to think hard about their answers. That perhaps deliberation comes effortlessly to them, which debunks the cognitive miser account. There was also an increase in judgement latencies under conflicting scenarios, which supports this idea. When faced with uncertainty, instead of simply responding with a fast, easy heuristic answer, they chose instead to spend longer thinking about their response, which would imply they are not cognitive misers at all. If people were simply cognitive misers, or they were not able to detect conflict, then they would respond to all statements (i.e., both those containing conflict and those without) in the same way and in the same amount of time. However, this was not the case. In fact, instead of people being “cognitive misers’ they appeared to be “effortless deliberators”. This interpretation of the data lends support for the heuristic reflections account, as it shows that people want to think hard about their answers, but at the same time also show high rates of reasoning errors because the bias enters human processing in System 2 (not System 1 as previously believed). If this is indeed the case, then System 2 reasoning might not be as cognitively taxing as previously believed, but this should be explored further in future research. See Table 6. 2 for an overview of the judgement latencies across the five experiments.

Table 6. 2

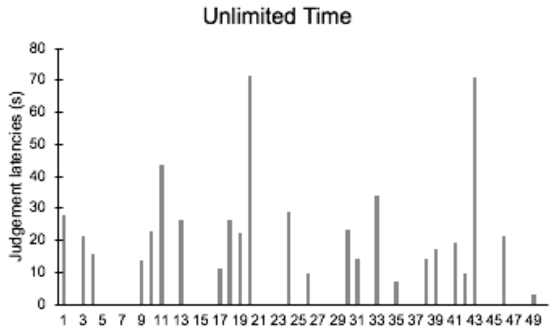
Overview of the Means, Standard Deviations, p-values and Effect Sizes of judgement latencies as a function of response type across the program of research. Note, Means and Standard Deviations are expressed in seconds

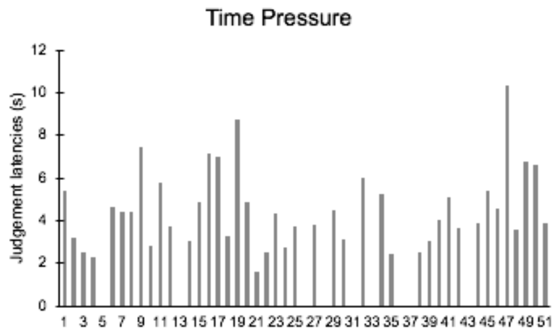
Experiment		*M* (*SD*)		*p*	Effect size
	No conflict	Conflict-Heuristic	Conflict-Logical		
1	9.97 (4.43)	10.90 (6.04)	14.20 (9.47)	< .001	$\eta_p^2 = .12$
2	12.31 (3.35)	12.05 (3.31)	12.57 (3.76)	.321	$\eta_p^2 = .02$
3	11.77 (5.71)	13.24 (6.94)	50.13 (36.19)	< .001	$\eta_p^2 = .63$
4	14.37 (7.01)	15.93 (8.10)	61.58 (49.25)	< .001	$\eta_p^2 = .50$
5	4.35 (2.40)	5.51 (3.04)	4.87 (3.89)	.860	$\eta_p^2 = .01$

When viewing the judgement latency data in entirety across all five studies, it becomes clear in three out of five studies that it took significantly longer to respond logically to conflict statements compared to the other types of responses. How does one explain this finding with regards to the logical intuitions account, which proposes that logical responses originate from intuitive thinking, thus they come first and are fast? Indeed, this result seems to support the original, dominant two-systems view as it suggests that System 2 overrides and corrects System 1. In situations of conflict, correct responding is assumed to require correction of an intuitive System 1 response by slower and more deliberate System 2 processing (Kahneman, 2011; Kahneman & Frederick, 2005). Thus, one needs to employ the slower deliberative thinking to reach the logical answer, which explains the long judgement latencies (Stanovich & West, 2000; Evans, 2008; Kahneman, 2011). In three of the experiments (excluding Experiment 2 and 5), when judgements latencies were grouped into type of response (e.g., No-Conflict, Conflict-Heuristic and Conflict-Logical), the logical judgements made when conflict was present in the statements took significantly longer to make compared to the heuristic judgements when conflict was present, or the judgements when conflict was absent from the statements. These results support the parallel-competitive account of reasoning, as they show that conflict is detected by participants. They took longer to respond heuristically to conflict statements compared to no conflict statements. Plus, the heuristic answer was given quickly compared to the logical answer when conflict was present, which suggests that logical judgements require the intervention of slower, more conscious deliberation to override the potent, fast heuristic response.

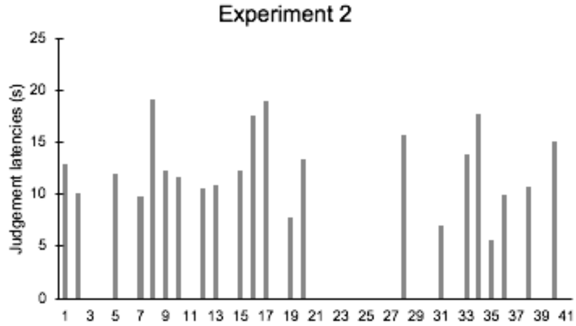
This is convincing evidence for the original parallel-competitive model of reasoning; however, realising that these judgement latencies are grouped together across manipulation conditions I decided to investigate them further as many of the experiments yielded significant interactions between the experimental manipulation and the judgement type. In doing so, I wanted to establish whether there were both fast and slow logical judgements being made on the conflict trials. A simple examination of the participants' judgement latencies when they answered conflict statements in line with logical considerations revealed that there were in fact both fast as well as slow logical judgements being made (see Figure 6. 1). This observation brings the parallel-competitive account into question, as it appears that sometimes logical responses can be arrived at very quickly, while other times they do take much longer. These incredibly fast logical answers (e.g., given in a matter of seconds) provides support for the logical intuitions account as these logical answers must be available first (i.e., as a System 1 response and before any influence from System 2), but also suggests that perhaps people have access to both fast and intuitive as well as slow and deliberate logical reasoning. This counteracts the traditional parallel-competitive account, which states that logical reasoning stems from slower, more effortful deliberation; however, it does not rule out a parallel model completely. Due to people's undeniable ability to detect conflict underlying logical and heuristic considerations, it is highly probable that they do rely on a parallel processing model of reasoning. The same way that people could have access to both logical intuitions and heuristic intuitions (e.g., logical intuitionist account), they may also have access to logical deliberations and heuristic deliberations (e.g., heuristic reflections account). Consider this possibility: the first response points towards a logical answer, but people being the opposite of cognitive misers in fact do want to think hard about their answers, thus they employ deliberation in a natural and effortless manner. At this point, the heuristic influences reasoning and biases the judgement. It is possible for a person to override this heuristic response and replace it with the original logical response, but this takes serious cognitive effort. This would explain the high rates of heuristic responding across the tasks, and also the presence of both very short and very long logical response latencies.

Experiment 1

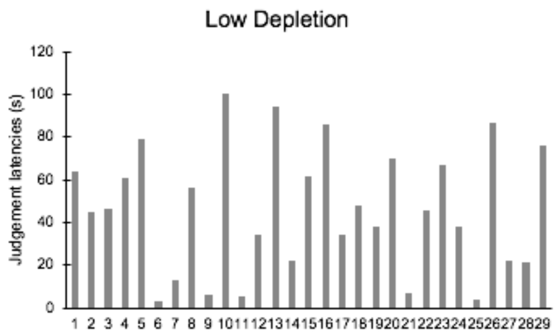


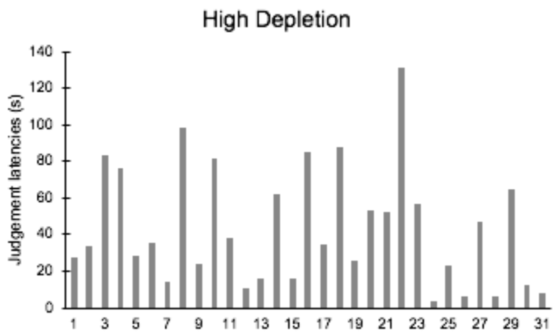


Experiment 2



Experiment 3





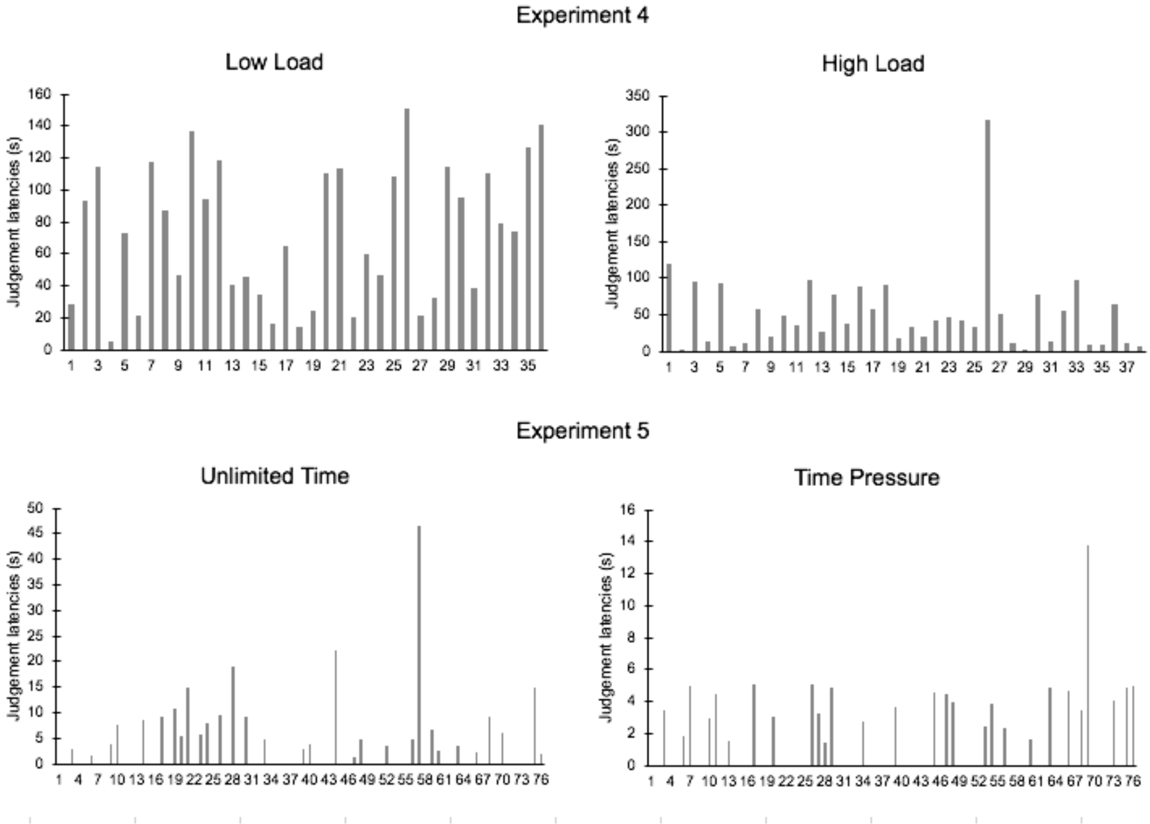


Figure 6. 1. Judgement latencies for the Logical-Conflict response type across the five experiments. *Note*. Latencies are expressed in seconds

In Experiments 2 and 5 there was little difference in response latencies between types of responses (see Table 6. 2). This did not necessarily mean that participants were unable to detect the conflict, as this was evident from their eye movement patterns and self-reported confidence scores. However, this conflict sensitivity did not materialise into their judgement latencies. In experiment 2, as previously mentioned, the choice in eye tracker may be the reason why. The Eyelink II is a serious looking piece of equipment with head mounted cameras and a chin and forehead bar. It might have been intimidating having this equipment strapped to one's head and as a result promoted deliberative thinking, which would explain the long judgement latencies across the three judgement types (e.g., 12 seconds). The results of experiment 5 were very unexpected and have been challenging to interpret because it was the only within-subjects study and also the only one to use a base rate neglect task, thus comparatively there is not much to work with. Time pressure, which had worked so effectively in previous studies, had no effect on participants' heuristic scores. While there was a significant difference between the unlimited time and time pressured conditions, such that trials answered under time pressure were also answered significantly faster; compared with the other four experiments this task was completed very rapidly. This might be due to the study's within-subjects design and also due to the difference in task (i.e., base rate neglect compared to conjunction fallacy). Hertwig and Gigerenzer (1999) discovered that the use of between-subjects groups or within-subjects groups made a difference to the rate of conjunction violations on the seven-letter word problem. Indeed, the between-subjects group violated the conjunction rule by estimating that more words end in 'ing' than in '_n_'. However, the within-subjects group, who saw both alternatives, answered the questions correctly and only 26% committed the conjunction fallacy. A similar discovery was made by Birnbaum and Mellers (1983) who found that the base rate fallacy also led to different conclusions depending on whether a within-subjects or between-subjects design was used. They found that in between-subjects comparisons the base rates were not given attention, thus base rate neglect was prevalent. However, in within-subjects comparisons participants used the base rates and the fallacy disappeared (Varey, Mellers, & Birnbaum, 1990). These authors cautioned against drawing inferences from between-subjects comparisons of judgement because these comparisons lead to paradoxical conclusions. In other words, there is a confound formed between the stimulus and the context by allowing the stimulus to evoke its own context. However, Tversky and Kahneman (1983) used both within- and between-subjects designs to test the Linda problem. In the between-subjects design participants either judged the constituent events (i.e., T, F) or their conjunction (i.e., T&F), not both, as in the within-subjects design. In both designs, the probability of T&F was ranked higher than the probability of T. The within-subjects design of Experiment 5 seems to have sped up the overall time taken to complete the trials. Although half of the trials were answered under time pressure and the other half answered with unlimited time; both of these conditions were completed very hastily compared to the CCPJ task across the first four experiments. Compared to the results of Experiment 1, the limited time conditions seem to be on a par with one another ($M_{\_E1\_limited}$ = 4.38s, $M_{\_E5\_limited}$ = 3.60s); however, the time pressured conditions vary greatly ($M_{\_E1\_unlimited}$ = 19.00s, $M_{\_E5\_unlimited}$ = 6.12s). Perhaps, informing participants before the study that they would be making judgements both under time pressure and without time pressure somehow urged them to answer faster than they would normally have. However, the base rate neglect task does differ to the conjunction fallacy task, and it might be plausible that participants interpreted the statements and made their judgements faster for the base rates compared to the conjunctions. However, these are simply speculations that need to be tested further through replication studies.

Methodological Observations

This research project was a learning experience from start to finish and is by no means a perfect piece of work. There are many potential limitations that will be discussed in more detail in this section; however, perhaps the most obvious and serious of the limitations in this body of research lies within the methodology. To be more precise, the CCPJ task used across the five experiments. This task was designed to be a forced-response task that contained only two choices, or responses. Participants were expected to either accept a statement as being true, or reject it as being false, based on reading a previous piece of information. Following the work of Evans and Curtis-Holmes (2005), this design attempted to disentangle the cognitive underpinnings of social judgements by allowing the researcher to assess whether the response came from heuristic or logical considerations. However, what this design did not allow for was the identification of guessing or ruling out the possibility that a participant was simply choosing the only other alternative available. In other words, a participant might reject a statement as false, but that does not mean he believes it to be true. Yet he has to choose true as it is the only other remaining option. Future research should develop and utilise a methodology better suited to the purpose of the study. This might include more responses to choose from. For example, Travers, Rolison and Feeney (2016) designed a task where each problem was presented in a 4-option multiple choice format. For the conflict items, the possible responses were the correct option, the incorrect option, and two incorrect foil options. For the no-conflict problems the correct intuitive option was presented with three incorrect foils.

An additional limitation to the current methodology is the way that the statements are answered (i.e., accept/reject). Both the CCPJ and the base rate neglect task use 'true or false' as a means of responding. Throughout this research program it has been assumed that accepting a statement as true is the equivalent of rejecting a statement as false. Both types of responses (e.g., true and false) were used to control for answer type. However, they may not simply mean the opposite of each other and might be interpreted differently from one another. To test for this potential confound, I calculated whether there was an interaction between conflict and response type using the control group from Experiment 1 (i.e., unlimited time condition). The amount of times each person accepted or rejected a statement when conflict was present and when conflict was absent in the statements was calculated. There was a significant main effect for type of response; participants rejected more statements overall than they accepted ($M_{\_accept}$ = 3.71, SD = 0.77, 95% CI [3.49, 3.93], $M_{\_reject}$ = 4.29, SD = 0.77, 95% CI [4.07, 4.51], $F(1, 49) = 7.09$, $p = .01$, $\eta_p^2 = .13$), but there was no interaction between conflict and type of response, $F(1, 49) = 0.32$, $p = .574$, $\eta_p^2 = .01$. This suggests that there is no difference between accepting a statement as true or rejecting a statement as false. Consequently, one can assume that participants were able to accept statements just as easily as they were to reject statements.

Another potential limitation to this body of research is the fact that it relies heavily on judgement latencies. Krajbich et al. (2015) show that using the reverse inference can be dangerous. The reverse inference is the use of behavioural or biological measures to infer mental function. It can be problematic as it does not consider other sources of variability in the data, for example discriminability of the choice options. Many researchers have used the assumption that decisions derived at through intuitive processes should tend to have shorter reaction times than those derived from deliberative processes (Kahneman, 2011; Evans & Stanovich, 2013). More recently, certain researchers have worked backwards from this to reason that fast decisions are intuitive (Stupple, Ball, Evans & Kamal-Smith, 2011; De Neys & Glumicic, 2008). However, there are many pitfalls associated with making reverse inferences, and this argument can also be applied to reaction time durations. There is "a key distinction between the prediction that an automatic process will occur faster than more deliberative computations, and the classification of a choice as intuitive or automatic because it happens more quickly" (Krajbich, et al., 2015, p. 2). These authors caution strongly against using reaction time differences as evidence favouring dual-process accounts as reaction time in a choice task depends critically on how different the decision maker finds the options available. For example, suppose a researcher designs a study on preferences between beer and wine, where subjects are required to make multiple decisions between different beer-wine pairings. The experimenter has to decide which beers and wines to include in her experiment. Depending on which brands she chooses, she might find that beers are generally preferred to the wines. While another experimenter running an identical experiment on the same population, might find that the wines he selected are preferred by his participants. The problem arises when these two researchers compare their reaction time results. The first researcher might find that the preference for beer amongst her subjects led to faster beer choices, while the second researcher found faster wine choices. Based on their results from these two seemingly identical studies they would reach opposite conclusions about whether people intuitively favour beer or wine in a fast, automatic way. This logic can be applied to any choice task. While the programme of research in this thesis does rely on reaction times, it does not use the reverse inference. What it does instead is compares latencies between judgements made under conflict and those without conflict and compares latencies between intuitive and logical assessments while under conflict. Although this body of work does state that logical intuitions are faster and come before heuristic deliberations, it does not put a timestamp on the ’fast’ or ‘slow’ judgements. What it attempts to do is to assess the timeline of judgement making by observing the order in which different types of responses are given (i.e., logical intuitions come first, followed by heuristic deliberations, and then possibly followed by logical deliberations). Additionally, this task might not be as vulnerable to mixed interpretations as the beer-wine study mentioned above. The Linda problem was designed to entice a heuristic response through the use of the representativeness heuristic. It was designed to do this by combining a highly likely scenario with a highly unlikely scenario, which makes the task easy to interpret as well as easy to commit the conjunction fallacy. Thus, when adapting Villejoubert’s (2009) study into the CCPJ task used in this thesis I first piloted the descriptions to ensure that people were surprised or shocked at one of the constituent events, while the other one seemed perfectly normal. The pilot study ensured that most people felt the same way about the fictitious characters used in the task. This would have standardised their answers to some extent and avoided individual preferences such as experienced in the beer-wine experiment. The design of the CCPJ task also illuminated the difference between a heuristic and logical assessment, such that it was possible to see which assessment had been used to judge the conflict statements. As previously mentioned, it is problematic to classify a choice as intuitive or automatic because it happens more quickly (Krajbich, et al., 2015). However, the CCPJ design now allows one to measure and compare the latencies between heuristic and logical judgements made under conflict (e.g., R/I and U/L statements); thus, allowing one to assess which comes first, or is more automatic.

Another major limitation to the study is the fact that for three out of a possible four conjunction fallacy experiments, the same control data was used. This is problematic because those control results have not been replicated and yet much of the evidence depends on comparisons with this data. The strength of the evidence in this body of research would have been much higher if new control data had been collected for at least one of Experiment 3 or 4.

A final limitation to this program of research would be to question whether or not I was in fact recording the first intuitive response. As with simple tasks such as Modus Ponens (e.g., if A then B; A; Therefore, B), the processing needed to arrive at the correct response is so basic that it is in fact automatized and assimilated as a System 1 response. With this in mind, could the correct intuitive answer possibly be attributed to the easiness of the task and not to logical intuitions (i.e., the automatization of laws of logic). I would suggest that the answer here is no. This is not an easy task. The task was designed to create stereotypes that contrasted greatly across categories. Each description was designed so that the person presented appeared both as a highly representative member of a stereotypical category and highly unrepresentative of an atypical category, without openly mentioning that this individual belonged to either category. The point of designing the descriptions in this fashion was to increase the conflict between intuitive and deliberative assessments, thus increasing the difficulty of the task. I would suggest that this CCPJ task, along with the base-rate neglect task, are difficult enough to warrant deliberation, which is why there is such a high rate of fallacies committed because the logical intuition is overridden by the heuristic deliberation. As I have discovered across this research program, people are effortless deliberators. They want to think hard about their answers. Perhaps if the task was simpler, they would be able to answer using more automatic responses, and thus we would see fewer fallacies committed. However, this is an assumption that would need to be tested further. Critics might also query whether the procedures used to guarantee a first intuitive response were stringent enough. A variety of methods were employed across the five experiments to encourage intuitive first responses, including extreme time pressure, cognitive load, and cognitive depletion. If as assumed, System 2 is heavily demanding and draining on cognitive resources, then by depriving participants of these resources we eradicate System 2 as much as possible during the initial response phase. Across the four experiments that used manipulations to decrease conjunction fallacies (i.e., increase logical judgements) three of them proved successful in significantly lowering the rates of conjunction fallacies committed by participants. See Table 6. 3 for more detail. Extreme time pressure, cognitive depletion and cognitive load (as between subjects designs) all significantly increased the logicality of people's conjunction probability judgements. Unfortunately, time pressure as a within subjects design and implemented on a base rate task had no effect on lowering base rate neglect. The purpose of the manipulations was to encourage System 1 thinking and eradicate System 2 thinking as much as possible. The pattern in the data suggests that these manipulations were stringent enough because they produced responses significantly different to the control conditions. The fact that these manipulations resulted in responses that were higher in logicality supports the notion of logical intuitions.

Table 6. 3

Means, Standard Deviations, p-values and Effect Sizes of heuristic scores as a function of task manipulation across the program of research. Note, Means and Standard Deviations are expressed as percentages

Experiment	Manipulation conditions	*M* (*SD*)	*p*	η_p^2
1	Unlimited time	88.63 (11.41)	< .001	.16
	Time pressure	76.72 (15.42)		
3	No depletion	88.52 (11.51)	.017	.06
	Low depletion	83.33 (13.42)		
	High depletion	80.96 (13.84)		
4	No load	88.28 (11.50)	.003	.08
	Low load	84.38 (12.69)		
	High load	78.86 (14.89)		
5	Unlimited time	88.80 (15.62)	.761	.001
	Time pressure	89.38 (14.25)		

Conflict being present in the statements as well as promoting intuitive thinking through various manipulations both resulted in more logical probability judgements. However, some might argue that these findings are simply due to a practice or learned effect. Experiment 1 took this into account by analysing the data across blocks. The task was designed so that the sixteen trials were presented over four blocks with four trials in each block. The trials within each block were randomised, and the order in which the blocks were presented was also randomised. Two different ordered versions of the task were used to minimise confounding variables. The order of the trials in version 1 was reversed for version 2 (e.g., an R&L trial in version 1 became a U&I trial in version 2). The results showed that there was no effect of blocks on either conflict sensitivity, or time pressure conditions, and there were no interactions found between these variables, which suggests that the order of the trials presented in each block had no effect on the logicality of the judgements produced. In other words, participants did not become more sensitive to the conflict between logical and heuristic considerations as they completed more trials. Thus, there was no evidence of practice or learned effects. The same CCPJ task was administered across the five experiments, which suggests there are no learned or practice effects in any of the studies within this program of research.

Concluding Remarks

This thesis contributes to the ever-expanding support for the existence of logical intuitions (De Neys, 2012; Villejoubert, 2009; Handley & Trippas, 2015; Trippas, Handley, Verde, & Morsanyi, 2016; Franssens & De Neys, 2009; Morsanyi & Handley, 2012; Pennycook, Trippas, Handley, & Thompson, 2013). It also contributes to the body of evidence that supports conflict detection. Recently, there is a burgeoning movement away from the lax monitoring perspective of the dominant two-system view, towards a belief that people are readily able to detect when deliberative and intuitive considerations are in conflict with one another (De Neys, Cromheeke & Osman, 2011; Pennycook, Fugelsang & Koehler, 2012; De Neys & Glumicic, 2008; De Neys, Vartanian, & Goel, 2008). However, the most important contribution to knowledge that this thesis provides is to propose a new explanation for the role that intuition plays in probability judgements, and in doing so, provides a different perspective on how dual processes of reasoning operate. Traditional dual process theories state that the intuitive heuristic response comes first but must be overridden by logical deliberation in order to avoid making an error. However, this program of research has provided evidence to support the idea that it is in fact the logical-intuitive answer that comes first, and this is followed by a heuristic-reflective answer when a person deliberates. It is entirely possible that people have learned the laws of probability through simply existing in the world. We might not have been formally taught these laws of statistics but have learned them implicitly by interacting with our environments (e.g., Saxe, 1988). This explains how these tenets of logic are possible as our first, intuitive response. However, it is when we start to deliberate and overthink our first response that the heuristic bias corrupts our judgement (i.e., prior beliefs and experiences taint our initial logical response). This program of research shows in fact that people are fond of deliberating. They want to think hard about their answer before making a judgement. In fact, it was extremely difficult to encourage people to respond using their intuitive System 1 thinking, especially under time pressure in Experiment 1. Increasingly strict measures had to be put in place to ensure participants would respond within 8 seconds. I piloted several methods, including a countdown clock, and a vocal command to answer immediately; however, these did not seem to portray the urgency that was necessary. Eventually, I discovered that by instructing participants that they would lose money each time they took too long to respond was quite motivating to the participants and this procedure was implemented in the final task design. The conjunction fallacy is extremely potent and once it has overridden the logical intuitive response it seems to take a great amount of cognitive energy to inhibit it and revert back to the initial logical answer. This is evidently difficult to do (as observed by the extended judgement latencies in the conflict-logical responses), but not entirely impossible as some participants were able to give the correct, logical responses to the conflict statements. The majority of participants; however, did still commit errors. Across this program of research, it was circa 80% which is in line with the original Linda problem (Tversky & Kahneman, 1983). This high rate of biased responding could be attributed to the fact that people are "effortless deliberators", in other words, they want to think hard about their answers and tend to employ this more effortful style of thinking very naturally (i.e., the opposite of a cognitive miser). People naturally want to think hard about their answers before making a judgement. Hence, the high rate of errors because the bias resides in deliberative System 2 thinking. It is invoked by reflecting on previously acquired knowledge and experiences. This is supported by the findings in Experiments 1, 3 and 4. Individuals who spent longer thinking about their answers also made more judgements in line with heuristic considerations. While individuals who were severely cognitively depleted, or cognitively overloaded (i.e., System 2 was suppressed) answered more in line with logical considerations compared to those in a control group. The findings across this research program have implications for dual process theories. They question the original two-systems accounts, both the serial and the parallel models (Evans, 2008; Stanovich & West, 2000; Sloman, 1996; Kahneman, 2003, 2011), which state that intuition will most often lead to heuristic, biased reasoning, but deliberation will most often lead to logical reasoning; offering instead an alternative explanation as to the role that intuition and deliberation plays in probability judgements. The findings are still aligned with a parallel account of reasoning as there was considerable evidence found for conflict sensitivity, hence the two systems ought to be running alongside one another in order for people to detect that they are producing differing responses. However, instead of the two systems running in parallel from the start (i.e., the traditional parallel-competitive model), System 2 might only activate at some point after System 1. System 2 is easily activated, as people are "effortless deliberators", so the trick is in fact finding ways to delay the activation of System 2 so that people respond in line with their initial logical answer. Some of the findings from this body of research are also aligned with the logical intuition account (De Neys, 2012), as far as supporting the idea that there are logical intuitions that come first; however, they differ in the detail of where the heuristic bias comes into play. The logical intuitionist account suggests that there are logical intuitions activated in parallel with heuristic intuitions right from the start; thus, System 1 thinking can produce biased or logical answers. While the heuristic reflective account suggests that the logical answer comes first. Logic is embedded in System

1 thinking, but when people switch to System 2 deliberation, that is where the bias is found, and that is when they make errors. The theory resulting from this programme of research, however, may propose that the logical answer comes first. It is a fast, automatic System 1 response derived from the implicit learning of logical laws by existing in and interacting with our environment. However, System 2 deliberation is easily employed by people, thus happens very naturally and is probably very difficult to avoid. It contains biased heuristics from previous knowledge and prior beliefs which easily override or inhibit the initial logical response. From this point on, the two Systems are running in parallel with each other, which can result in feelings of conflict between the intuitive and deliberate considerations. The heuristic-reflective response presents convincing evidence, which could explain the high rates of heuristic responding seen in the CCPJ task. Those participants who solved the problems correctly (e.g., logically) after spending considerable time on each problem probably needed an immense amount of cognitive resources and deliberation to inhibit the deliberate-heuristic response and replace it with a deliberate-logical one. In other words, it might be possible that people have access to both intuitive-logical and deliberate-logical judgements. Thus the sequence of events would unfold as such: logical intutions come first, but are overridden by heuristic reflections, which might in turn be overridden by logical deliberations. There is the possibility that logical deliberations run parallel to heuristic deliberations, which allows or conflict detection between the two, and also means that if enough cognitive energy is invested in the judgement, people would be able to override the heuristic reflective response and replace it with the original logical response, thus making a logical-deliberative judgement. See Figure 6. 2 for an overview of the different dual process models displaying the relations between the intuitive and deliberative system. However, it is very important to state that at this point it is purely a theoretical claim that attempts to explain the roles that intuition and deliberation play when people make social judgements. In fact, this theory cannot be applied beyond the scope of conjunction and base rate judgements at the present

time. This theory was derived by attempting to consolidate the data found across these five experiments with the existing dual process theories of reasoning. It is crucial that further research is conducted, using valid and reliable methodologies, to test these theoretical claims. The dominant two-system view has been applied to many forms of complex cognitive activities (Darlow & Sloman, 2010; Reyna & Brainerd, 1995), thus the findings from this research program promise to have wide ranging implications for our understanding of how intuition and deliberation interact when we engage in such activities.

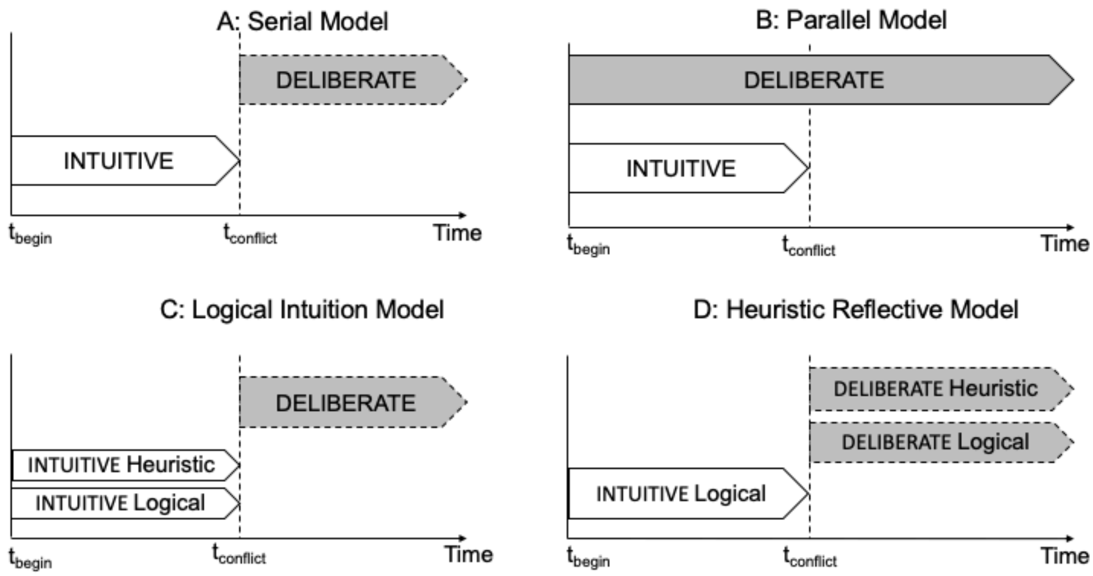


Figure 6. 2. Four different theoretical models of the relation between the intuitive and deliberative systems. Deliberate processing is represented by grey bars and intuitive processing by white bars. The horizontal axis represents the flow of time. Dashed lines represent the optional nature of the triggered deliberative processing (adapted from De Neys, 2012). The heuristic reflective model is originally proposed here (but see also Handley & Trippas, 2015)

Future research

This thesis examined the effect of limiting people's deliberation through the use of time pressure, ego depletion and cognitive load, and in doing so encouraged intuitive responding. Future research might consider ways to increase the potency of people's logical intuitions. In other words, instead of constraining deliberation, one might explore ways to increase people's trust in their logical intuitions. One example could be to prime intuitive responding before completing the CCPJ task. Bakhti (2018) used the scrambled sentence task to prime individuals before completing conjunction fallacy tasks and found that individuals who had been religiously primed committed more fallacies than those who had been reflectively primed. Additionally, one might look at adapting or developing a new task to test the interplay between intuition and logical assessments when they are on equal playing fields. It is important to improve on the current CCPJ task, and this new task should use non-binary response options to allow the researcher to gain more insight as to the mechanisms of reasoning when making these types of judgements. Dual process theorists suggest that heuristic processes are associative (e.g., Sloman, 1996). "One very important characteristic of outputs from associative processes is that they vary in strength. Thus, the output of heuristic processes should vary in strength, and the ability of analytic processes to control such outputs should, in turn, vary with the strength of the heuristic process" (Crisp & Feeney, 2009, p. 2323). Indeed, the authors found that the rate of conjunction fallacies was higher for strongly related conjunctions than for weakly related ones, thus replicating the findings of Tversky and Kahneman (1983). However, they also found that strongly related conjunctions cued a much stronger heuristic output than the weak conjunctions, and the strength of this output determined the ease with which the analytical system inhibited the heuristic response. The authors suggest that "causal knowledge has a crucial influence on probabilistic reasoning" (Crisp & Feeney, 2009, p. 2331). The original Linda problem was designed to produce high rates of the conjunction fallacy, which is exactly what it does (e.g., circa 80% of all participants commit the fallacy; Tversky & Kahneman, 1983). It has strong causal links between the events and was developed to compare a highly likely category with a highly unlikely one. The representativeness heuristic, which is evoked through the description, is what makes the conjunction so compelling in conflict scenarios. It would however, be interesting to see what happens in a task where intuition and logic are “pulled at” with equal strength. Something similar perhaps to the unrelated scenarios as described in Crisp and Feeney (2009; see Experiment 1). In all likelihood, people would probably still naturally want to deliberate as this has been evidenced throughout this research program. People engage unprompted and easily in deliberation, and it proved difficult at times to get them to respond using their intuitive thought only. Thus, one might expect lower rates of fallacies rather than a complete elimination. Another future research avenue might involve examining more closely the time it takes for the logical intuition to be overridden by the deliberative heuristic. There might be a “sweet spot”, or a range in time, when people are able to apply normative laws to their judgements (whether they are aware of doing so or not). This time frame may well differ across tasks. It did in this research program as seen between the conjunction fallacy and base rate neglect tasks, although this may be also due to the strength of the heuristic response. Identifying the moment at which heuristic deliberation “contaminates” logical intuition may provide an effective mean to minimise people’s propensity to commit the conjunction fallacy.

References

Abelson, R.P., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology (Volume 1): Theory and method* (pp. 231– 309). New York: Random House.

Albertazzi, L. (2006). *Visual thought.* Amsterdam/Philadelphia: John Benjamins.

Alos-Ferrer, C., & Hugelschafer, S. (2012). Faith in intuition and behavioural biases. *Journal of Economic Behaviour & Organisation, 84,* 182-192. doi: 10.1016/j.jebo.2012.08.004

Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology, 45*, 11-19. https://doi.org/10.1111/j.1469-8986.2007.00609.x

Arber, M. M., Ireland, M. J., Feger, R., Marrington, J., Tehan, J., and Tehan, G. (2017). Ego depletion in real-time: an examination of the sequential-task paradigm. *Frontiers in Psychology, 8,* 1-12. doi: 10.3389/fpsyg.2017.01672

Ashcraft, M. H., & Kirk, E. P. (2001). The relationship among working memory, math anxiety, and performance. *Journal of Experiemntal Psychology: General, 130,* 224-237. doi: 10.1037/0096-3445.130.2.224

Baddeley, A. (1992). Working Memory. *Science, 225,* 556-559. doi: 10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4,* 417-423.

Baddeley, A. D. (2007). Working memory, thought and action. Oxford, UK: Oxford University Press. Retrieved from http://dx.doi.org/10.1093/acprof:oso/9780198528012.001.0001

Bago, B., & De Neys, W. (2017). Fast logic? Examining the time course assumption of dual process theory. *Cognition, 158,* 90-109. doi: 10.1016/j.cognition.2016.10.014

Bago, B., & De Neys, W. (in press). The smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*. doi: 10.1080/13546783.2018.1507949

Bakhti, R. (2018). Religious versus reflective priming and susceptibility to the conjunction fallacy. *Applied Cognitive Psychology, 32,* 186-191. doi: 10.1002/acp.3394

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

Bamberger, P. A. & Bacharach, S. B. (2006). Abusive supervision and subordinate problem drinking: Taking resistance, stress and subordinate personality into account. *Human Relations, 59,* 723-752. doi: 10.1177/0018726706066852

Bar-Hillel, M. (1980). The base rate fallacy in probability judgements. *Acta Psychologica, 44,* 211-233.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgements. *Journal of Personality and Social Psychology, 65,* 1119-1131. doi: 10.1037/0022-3514.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioural and Brain Sciences, 30,* 241-256.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71,* 230-244. doi: 10.1037/0022-3514.71.2.230

Bargh, J. A., Lee-Chai, A., Barndollar, K., Gollwitzer, P. M., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioural Goals. *Journal of Social Psychology, 81,* 1014-1027. doi: 10.1037/0022-3514.81.6.1014

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,* 553-573. doi: 10.1037/0033-2909.130.4.553

Baumeister, R. F., and Vohs, K. D. (2016). Misguided effort with elusive implications. *Perspectives on Psychological Science, 11,* 574–575. doi: 10.1177/1745691616652878

Baumeister, R.F., Bratslavsky, E., Muraven, M., & Tice, D.M. (1998). Ego-depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74,* 1252-1265.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 12-23. Doi: 10.1037/0096-1523.14.1.12

Birnbaum, M. H. & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45*, 792-804.

Böckenholt, U. (2012). The coginitve-miser response model: Testing for intuitive and deliberative reasoning. *Psychometrika 77,*388-399. doi: 10.1007/S11336-012-9251-Y

Bonini, N., Tentori, K., & Osherson, D. (2004). A different conjunction fallacy. *Mind & Language, 19,* 199–210. doi: 10.111/j.1468-0017.2004.00254.x

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*, 186–196.

Brotherton, R. & French, C. C. (2014). Belief in conspiracy theories and susceptibility to the conjunction fallacy. *Applied Cognitive Psychology, 28,* 238-248. doi: 10.1002/acp.2995

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

Camerer, C. (1995) “Individual Decision-making,” *The Handbook of Experimental Economics,* Eds. John Kagel and Alvin Roth. Princeton: Princeton University Press.

Campos, F., Lincoln, D., Neves, M., Correia, W., & Soares, M. (2013). The conjunction fallacy and its impact in user’s data acquisition process. In: Marcus A. (eds) Design, User Experience, and Usability. Design Philosophy, Methods, and Tools. DUXU 2013. Lecture Notes in Computer Science, vol 8012. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-39229-0_23

Carter, E. C., Kofler, L. M., Forster, D. E., and McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology General. 144,* 796–815. doi: 10.1037/xge0000083

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated? *Frontiers in Psychology, 5,* 1-11. doi: 10.3389/fpsyg.2014.00823

Charness, G., Karni, E., & Levin, D. (2010). On the conjunction fallacy in probability judgement: New experimental evidence regarding Linda. *Games and Economic Behavior, 68,* 551-556. doi: 10.1016/j.geb.2009.09.003

Chen, Y, & Sun, Y. (2011). Age differences in financial decision-making: Using simple heuristics. *Educational Gerontology, 29,* 627-635. Doi: 10.1080/713844418

Costa, A., & Kallick, B. (2000). Discovering and exploring habits of mind. Alexandria, VA: Association for Supervision and Curriculum Development.

Costello, F.J. (2009). How probability theory explains the conjunction fallacy. *Journal of Behavioural Decision Making, 22,* 213-234. doi: 10/1037/a0037010

Costello, F. & Watts, P. (2017). Explaining high conjunction fallacy rates: The probability theory plus noise account. *Journal of Behavioral Decision Making, 30,* 304-321. doi: 10.1002/bdm.1936

Crisp, A. K., & Feeney, A. (2009). Causal conjunction fallacies: the roles of causal strength and mental resources. *Quarterly Journal of Experimental Psychology, 62,* 2320-37. doi: 10.1080/17470210902783638

Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 382-392. doi: 10.1002/wcs.34

Dearden, T. E. (2018). The conjunction fallacy in profiles of victims of homicide. *Journal of Investigative Psychology and Offender Profiling, 15,* 187-199. doi: 10.1002/jip.1502

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic – an fMRI study. *Cognitive Brain Research, 18,* 76-88. doi: 10.1016/j.cogbrainres.2003.09.005

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgement. *Journal of Personality and Social Psychology, 66,* 819–829.

De Neys, W. (2006a). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428-433. doi: 10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*, 1070-1100. doi: 10.1080/02724980543000123

De Neys, W. (2009). Beyond response output: More logical than we think. *Behavioral and Brain Sciences, 32,* 87-88. doi: 10.1017/S0140525X09000326

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

De Neys, W. (2013). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking and Reasoning, 20,* 169–187. doi:10.1080/13546783. 2013.854725

De Neys, W. (2015). Heuristic bias and conflict detection during thinking. *Psychology of Learning and Motivation, 62,* 1-32. doi: 10.1016/bs.plm.2014.09.001

De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences, 17,* 172-178. doi: 10.1016/j.tics.2013.02.001

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. PLoS ONE, 6, 1-10. doi: 10.1371/ journal.pone.0015954

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248-1299. doi:10.1016/j.cognition.2007.06.002

De Neys, W., & Feremans, V. (2013). Development of heuristic bias detection in elementary school. *Developmental Psychology, 49,* 258–69.

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

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience, 10,* 208–216. doi: 10.3758/CABN.10.2.208

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54,* 128-133. https://doi.org/10.1027/1618-3169.54.2.128

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

Denis, M., Logie, R., & Cornoldo, C. (2012). The processing of visuo-spatial information: Neuropsychological and neuroimaging investigations. In *Imagery, Language and Visuo-Spatial Thinking* (pp. 81-102). Hove, US: Psychology Press.

Devine, P.G. & Moneith, M.J. (1999). Automaticity and Control in Stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology* (pp. 339-360). New York: The Guilford Press.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlate. *Journal of Personality and Social Psychology, 58,* 95-102.

Droll, J.A., & Hayhoe, M.M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1352-1365. Doi: 10.1037/0096-1523.33.6.1352.

Duffy, S. A, & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32*, 551-559.

Epstein, S. (1994). Intergration of the Cognitive and the Psychodynamic Unconscious. *American Psychologist, 49,* 709-724. doi: 10.1037//0003-066X.49.8.709.

Epstein, S., Pacini, R., Denes-Raj, V., & Heiser, H. (1996). Individual differences in intuitive – Experiential and analytical – Rational thinking styles. Journal of Personality and Social Psychology, 2, 390–405. Available from: https://www.researchgate.net/profile/Seymour_Epstein/publication/14436248_Individual_Differences_in_Intuitive-Experiential_and_Analytical-Rational_Thinking_Styles/links/5588bd7308aeb2994444a7bb/Individual-Differences-in-Intuitive-Experiential-and-Analytical-Rational-Thinking-Styles.pdf

Erceg, N. & Galić, Z. (2014). Overconfidence bias and conjunction fallacy in predicting outcomes of football matches. *Journal of Economic Psychology, 42,* 52-62. doi: 10.1016/j.joep.2013.12.003

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decsion conflict drives reaction times in social dilemmas. *Journal of Exerimental Psychology: General, 144,* 951-966. Doi: 10.1037/xge0000107

Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75,* 451–468.

Evans, J.St.B.T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454–459. Doi: 10.1016/j.tics.2003.08.012.

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

Evans, J. St. B. T. (2007). On the resolution of conflict in dual- process theories of reasoning. *Thinking & Reasoning, 13,* 321–329. doi: 10.1080/13546780601008825

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

Evans, J. St. B. T. (2009). How many dual process theories do we need: One, two or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford, UK: Oxford University Press.

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

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11,* 382-389. doi: 10.1080/13546780542000005

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

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223-241. doi: 10.1177/1745691612460685.

Featherstone, C. R., & Sturt, P. (2010). Because there was a cause for concern: An investigation into a word specific prediction account of the implicit-causality effect. *Quarterly Journal of Experimental Psychology, 63*, 3-15.

Feeney, A., Shafto, P., & Dunning, D. (2007). Who is susceptible to conjunction fallacies in category-based induction? *Psychonomic Bulletin & Review, 14,* 884-889.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123-129. doi: 10.1007/BF00309212

Fisk, J. E. (2002). Judgements under uncertainty: Representativeness or potential surprise? *British Journal of Psychology, 93*, 431-449.

Fisk, J. E., & Pidgeon, N. (1996). Component probabilities and the conjunction fallacy: Resolving signed summation and the low component model in a contingent approach. *Acta Psychologica, 94*, 1-20.

Fiske, S. T., & Taylor, S. E. (1991). *McGraw-Hill series in social psychology. Social cognition (2nd ed.).* New York, NY, England: Mcgraw-Hill Book Company.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15,* 105-128. doi: 10.1080/13546780802711185f

Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives, 19, 25–42.

Freud, S. (1961). The ego and the id. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol 19, pp. 12-66). London: Hogarth Press. (Original work published 1923).

Friedman, R. S., & Förster, J. (2000). The effects of approach and avoidance motor actions on the elements of creative insight. *Journal of Personality and Social Psychology, 79,* 477-492.

Gaissmaier, W., Fific, M. & Rieskamp, J. (2011). Analyzing response times to understand decision processes. In M. Schulte-Mecklenbeck, A. Kühberger & R. Ranyard (Eds.), A handbook of process tracing methods for decision research: A critical review and user's guide (pp. 141-162). New York and Hove: Psychology Press.

Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., et al. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology, 92*, 325–336.

Gerrard, M., Gibbons, F. X., Houlihan, A. E., Stock, M. L., & Pomery, E. A. (2008). A dual-process approach to health risk decision making: The prototype willingness model. *Developmental Review, 28,* 29-61. Doi: 10.1016/j.dr.2007.10.001.

Gidlof, K., Wallin, A., Dewhurst, R. & Holmqvist, K. (2013). Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research, 6*, 1-14. Doi: 10.16910/jemr.6.1.3.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129- 161). New York: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*, 592–596.

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology, 62*, 451-482. doi: 10.1146/annurev-psych-120709-145346

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

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Ecomonics, 4,* 125-146. doi: 10.1037/a0020692

Glöckner, A., & Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making, 24,* 71-98.

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgment and decision making. *Thinking & Reasoning, 16,* 1-25.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*, 546–573. Doi: 10.1177/1745691616652873.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136,* 495-525. Doi: 10.1037/a0019486

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 28-43. doi: 10.1037/a0021098

Handley, S. J., & Trippas, D. (2015). Dual processes, knowledge, and structure: A critical evaluation of the default interventionist account of biases in reasoning and judgement. *Psychology of Learning and Motivation, 62*, 33–56. doi: 10.1016/bs.plm.2014.09.002

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of and. *Cognition, 108,* 740-753.

Hertwig, R. & Gigerenzer, G. (1999). The "conjunction fallacy" revised: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12,* 275-305

Heuristic. (2019). Online Etymology Dictionary. Retrieved from https://www.etymonline.com/columns/post/links?utm_source=etymonline_footer&utm_medium=link_exchange

Hoppe, E.I., Kusterer, D.J. (2011). Behavioral biases and cognitive reflection. *Economics Letters 110,* 97–100.

Horstmann, N., Ahlgrimm, A., & Glöckner, A. (2009). How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes. *Judgement and Decision Making, 4,* 335-354.

Houdé, O., & Moutier, S. (2003). Judgement under uncertainty and conjunction fallacy inhibition training. *Thinking and Reasoning, 9,* 185-201. doi: 10.1080/13546780343000213.

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

Howarth, S., Handley, S., & Walsh, C. (2018). The logic sense: Exploring the role of executive functioning in belief and logic-based judgements. *Thinking and Reasoning,* 1-33. doi: 10.1080/13546783.2018.1523808

Howarth, S., Handley, S., & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief-logic conflict. *Memory and Cognition, 44,* 330-349. doi: 10.3758/s1342

Inzlicht, M., & Gutsell, J. N. (2007). Running on empty: Neural signals for self-control failure. *Psychological Science, 18,* 933–937.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion – Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21,* 1686-1693. Doi: 10.1177/0956797610384745.

Johnson, E.D., Tubau, E., & De Neys, W. (2016). The doubting system 1: Evidence for automatic substitution sensitivity. *Acta Psychologica, 164,* 56-64. doi: 10.1016/j.actpsy.2015.12.008

Kahneman, D. (1991). Judgment and decision making: A personal view. *Psychological Science, 2*, 142-145.

Kahneman, D. (2011). *Thinking, fast and slow*. London, UK: Penguin Group.

Kahneman, D. (2003). A perspective on judgement and choice: mapping bounded rationality. *The American Psychologist, 58,* 697-720.

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

Kahneman D., & Tversky A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology, 3,* 430-454. doi: 10.1007/978-94-010-2288-0_3

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251. doi: 10.1037/0033-295X.90.4.293

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103,* 582-591.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639.

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications, 6,* 1-9. doi: 10.1038/ncomms8455

Lu, Y. (2015). Is experiential-intuitive cognitive style more inclined to err on conjunction fallacy than analytical-rational cognitive style? *Frontiers in Psychology, 6,* 1-8. doi: 10.3389/fpsyg.2015.00085.

Lurquin, J. H., Michaelson, L. E., Barker, J. E., Gustavson, D. E., von Bastian, C. C., Carruth, N. P., & Miyake, A. (2016). No evidence of the ego-depletion effect across task characteristics and individual differences: a pre-registered study. *PLoS One, 11*, 1-20. doi: 10.1371/journal.pone.0147770

Martijn, C., Tenbült, P., Merckelbach, H., Dreezens, E., & de Vries, N.K. (2002). Getting a grip on ourselves: Challenging expectations about loss of energy after self-control. *Social Cognition, 20,* 441–460. Doi: 10.1521/soco.20.6.441.22978

Martin, L. E. & Potts, G. F. (2009). Impulsivity in decision-making: An event-related potential investigation. *Personality and Individual Differences, 46,* 303-308. doi: 10.1016/j.paid.2008.10.019

Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: lemonade, willpower, and expensive rule-based analysis. *Psychological Science, 19*, 255-60.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. doi: 10.1037/a0028085

Medin, D., Coley, J. D., Storms, G., & Hayes, B. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*, 517-532.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). DO frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269-275.

Mevel, K,, Poirel, N., Rossi, S., Cassotti, M., Simon, G., Houdé, O., &De Neys, W. (2015). Bias detection: Response confidence evidence for conflict detection sensitivity in the ratio bias task. *Journal of Cognitive Psychology, 27,* 227-237. doi: 10.1080/20445911.2014.986487

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers.* London: Sage.

Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640. Doi. 10.1037/0096-3445.130.4.621.

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry, 158,* 1783–1793.

Moller, A.C., Deci, E.L., & Ryan, R.M. (2006). Choice and ego- depletion: The moderating role of autonomy. *Personality and Social Psychology Bulletin, 32,* 1024–1036.

Moors, A., & de Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. Psychological Bulletin, 132, 297–326.

Moro, R. (2009). On the nature of the conjunction fallacy. *Synthese, 71,* 1-24. doi: 10.1007/s11229-008-9377-8

Morsanyi, K., & Handley, S. J. (2012). Logic feels so good—I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 596-616. http://dx.doi.org/10.1037/a0026099

Moutier, S., & Houdé, O. (2003). Judgement under uncertainty and conjunction fallacy inhibition training. *Thinking & Reasoning, 9,* 185-201. doi: 10.1080/13546780343000213

Muraven, M., Collins, R. L., & Nienhaus, K. (2002). Self-control and alcohol restraint. An initial application of the self-control strength model. *Psychology of Addictive Behaviors, 16,* 113–120.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: regulatory depletion patterns. *Journal of Personality and Social Psychology, 74,* 774–789. doi: 10.1037/0022-3514.74.3.774.

Muraven, M., & Slessareva, E. (2003). Mechanism of self-control failure: Motivation and limited resource. *Personality and Social Psychology Bulletin, 29*, 894–906. Doi: 10.1177/0146167203253209.

Myers, L., Downiw, S., Taylor, G., Marrington, J., Tehan, G., & Irelands, M. J. (2018). Understanding performance decrements in a letter-cancelling task: Overcoming habits or inhibition of reading. *Frontiers in Psychology, 9,* 1-15. Doi: 10.3389/fpsyg.2018.00711.

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2009). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition, 37,* 632–643.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioural biases. *Journal of Economic Behaviour & Organization, 72,* 147-152. doi: 10.1016/j.jebo.2009.04.018

O'Hagan, A., Buck, C. E., Daneshkhan, A., Eliser, J. R., Garthwaite, P. H., Jenkinson, D. J.,…Rakow, T. (2006). *Uncertain Judgements: Eliciting experts' probabilities.* West Sussex, UK: Wiley.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972–987. doi: 10.1037/0022-3514.76.6.972

Patton, J.H., Stanford, M.S., & Barratt, E.S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 6,* 768-774.

Pennycook, Fugelsang and Koehler (2012). Are we good at detecting conflict during reasoning? *Cognition, 124*, 101-106. doi: 10.1016/j.cognition.2012.04.004

Pennycook, G. & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. Psychonomic Bulletin & Review, 19, 528-534. Doi: 10.3758/s13423-012-0249-3.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2013). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 544-554. doi: 10.1037/a0034887

Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research Marketing, 16,* 1-16.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2015). Talker-Specific Generalization of Pragmatic Inferences based on Under- and Over-Informative Prenominal Adjective Use. *Frontiers in Psychology, 6,* 1-18. http://doi.org/10.3389/fpsyg.2015.02035

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174.

Recalde, M. P., Riedl, A. & Vesterlunde, L. (2014). Error prone inference from response time: The case of intuitive generosity. CESifo Working Paper Series No. 4987. Available at SSRN: http://ssrn.com/abstract=2507723 (Date of access: 17/01/2019)

Reeves, T. & Lockhart, R. S. (1993). Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General, 122,* 207-226.

Reisen, N., Hoffrage, U., & Mast, F.W. (2008). Identifying decision strategies in consumer choice situation. *Judgement and Decision Making, 3,* 641-658.

Reverberi, C., Pischedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta Psychologica, 139,* 244-253.doi:10.1016/j.actpsy.2011.09.011

Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review, 11,* 317-336.

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-trace theory. *Judgment and Decision Making, 7*, 332-359.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1-75. doi: 10.1016/1041-6080(95)90031-4

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences, 18*, 89–107. doi:10.1016/j.lindif.2007.03.011

Reyna, V.F., Lloyd, F.J., & Brainerd, C.J. (2003). Memory, development, and rationality: An integrative theory of judgement and decision- making. In S. Schneider & J. Shanteau (Eds.), Emerging perspectives on judgment and decision research (pp. 201–245). New York: Cambridge University Press.

Rieger, M. O. (2012). Why do investors buy bad financial products? Probability misestimation and preferences in financial investment decision. *Journal of Behavioral Finance, 13,* 108-118. doi: 10.1080/15427560.2012.680991

Riis, J. & Schwarz, N. (2003). Approaching and avoiding Linda: Motor signals influence the conjunction fallacy. *Social Cognition, 21,* 247-262.

Russo, J.E. (2011). Eye fixations as a process trace. In M. Schulte-Mecklenbeck, A. Kühberger & R. Ranyard (Eds.), A handbook of process tracing methods for decision research: A critical review and user's guide (pp. 43-64). New York and Hove: Psychology Press.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17,* 759–769.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763-776. doi: 10.1037/0012-1649.27.5.763

Saxe, G. (1988). The mathematics of child street vendors. *Child Development, 59*, 1415-1425.

Schmeichel, B. J., Harmon-Jones, C. & Harmon-Jones, E. (2010). Exercising self-control increases approach motivation. *Journal of Personality and Social Psychology, 99,* 162-173. doi: 10.1037/a0019797

Schmeichel, B. J., & Vos, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology, 96,* 770-782. doi: 10.1037/a0014635

Shiloh, S., Salton, E., & Sharabi, D. (2002). Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Personality and Individual Differences, 32,* 415-429.

Sides, A., Osherson, D., Bonini, N. & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory and Cognition, 30,* 191–198. doi: 10.3758/BF03195280

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology, 6,* 1-16. doi: 10.3389/fpsyg.2015.00532

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

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91,* 296-309. doi:10.1016/S0749-5978(03)00021-9

Spivey, M., Richardson, D.C., & Dale, R. (2009). Movements of eye and hand in language and cognition. In E. Morsella, J. Bargh, & P.M. Gollwitzer (Eds.), *The psychology of action* (Vol. 2, pp. 225-249). New York: Oxford University Press.

Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science, 25*, 1227-1234.

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

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate [Target article and commentaries]. *Behavioral and Brain Sciences, 23*, 645–726.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94,* 672-695. Doi: 0.1037/0022-3514.94.4.672

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31,* 103-118. doi: 10.1016/j.dr.2011.07.003

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking and Reasoning, 14,* 168-181. doi: 10.1080/13546780701739782

Stupple, E. J. N., Ball, L. J., Ellis, D. (2013). Matching bias in syllogistic reasoning: Evidence for a dual process account from response times and confidence ratings. *Thinking & Reasoning, 19*, 54-77. doi: 10.1080/13546783.2012.735622

Stupple, E. J. N., Gale, M., & Richmond, C. R. (2013). Working memory, cognitive miserliness and logic as predictors of performance on the Cognitive Reflection Test. *Proceedings of the Annual Meeting of the Cognitive Science society, 35*, 1396-1401. Austin, TX: Cognitive Science Society. Retrieved from https://escholarship.org/content/qt36989187/qt36989187.pdf

Teigen, K. H. (2004). Judgement by representativeness. In R. Pohl (Ed.), *Cognitive illusions: Handbook on fallacies and biases in thinking, judgement, and memory* (pp. 165-182). Hove, UK: Psychology Press.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science, 28*, 467-477.

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20,* 215-244. doi: 10.1080/13546783.2013.869763

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. Cognitive Psychology, 63, 107-140. doi: 0.1016/j.cogpsych.2011.06.001

Thurstone, L.L. (1927). *The nature of intelligence.* New York: Harcourt.

Tice, D.M., Baumeister, R.F., Shmueli, D., & Muraven, M. (2007). Restoring the self: Positive affect helps improve self-regulation following ego depletion. *Journal of Experimental Social Psychology, 43*, 379–384.

Tice, D. M., & Bratslavsky, E. (2000). Giving in to feel good: The place of emotion regulation in the context of general self-control. *Psychological Inquiry, 11,* 149-159. doi: 10.1207/S15327965PLI1103_03

Toplak, M. E., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making, 20*, 103-124. doi: 10.1002/bdm.544

Toplak, M. E., West, R.F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition, 39,* 1275-1289. Doi: 10.3758/s13421-011-0104-1.

Toyosawa, J., & Karasawa, K. (2004). Individual differences on judgment using the ratio-bias and the Linda problem: Adopting CEST and Japanese version of REI. *The Japanese Journal of Social Psychology 20,* 85–92.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition, 150,* 109-118. doi: 10.1016/j.cognition.2016.01.015

Tremoliere, B. & De Neys, W. (2014). When intuitions are helpful: Prior beliefs can support reasoning in the bat-and-ball problem. *Journal of Cognitive Psychology, 26,* 486-490. doi:10.1080/20445911.2014.899238

Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1448-1457. doi: 10.1037/xlm0000248

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

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

Tversky, A., & Kahneman, D. (1982). Judgement under uncertainty: heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky, *Judgement under uncertainty: Heuristics and biases* (pp. 3-20). Cambridge, UK: Cambridge University Press.

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

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 613-625.

Vartanian, O., Beatty, E. L., Smith, I., Blackler, K., Lam, Q., Forbes, S., & De Neys, W. (2018). The reflective mind: Examining individual differences in susceptibility to base rate neglect with fmri. *Journal od Cognitive Neuroscience, 30,* 1011-1022. doi: 10.1162/jocn_a_01264

van Raaij, F. W. (1977). Consumer information processing for different information structures and formats. *Advances in Consumer Research, 4*, 176–184.

Velichkovsky, B., Dornhoefer, S., Kopf, M., Helmert, J., & Joos, M. (2002). Change detection and occlusion modes in road-traffic scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour, 5,* 99-109.

Villejoubert, G. (2009). Are representativeness judgments automatic and rapid? The effect of time pressure on the conjunction fallacy. *Proceedings of the Annual Meeting of the Cognitive Science society, 30*, 2980–2985. Austin, TX: Cognitive Science Society. Retrieved from http://141.14.165.6/CogSci09/papers/655/paper655.pdf

Villejoubert, G. (2011). Rethinking the Role of Intuition in the Conjunction Fallacy. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.). *Perspectives on Thinking, Judging, and Decision Making* (pp. 109–119). Oslo: Universitetsforlaget.

Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science, 11*, 249–254.

Welsh, M. B., Burns, N. R., & Delfabbro, P. H. (2013). The Cognitive Reflection Test: how much more than numerical ability? *Proceedings of the Annual Meeting of the Cognitive Science society, 35*, 1587-1592. Retrieved from https://escholarship.org/uc/item/68n012fh

Whiteside, S. P. & Lynam, D. R. (2001). The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30,* 669-689. doi:10.1016/S0191-8869(00)00064-7

Wilson Van Voorhis, C.R., & Morgan, B.L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorial in Quantitative Methods for Psychology, 3,* 43-50.

Wojciechowski, B. W., & Pothos, E. M. (2018). Is there a conjunction fallacy in legal probabilistic decision making? *Frontiers in Psychology, 9*, 1-11. doi: 10.3389/fpsyg.2018.00391

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P.,…Wing, R. R. (2014). Failure to replicate depletion of self-control. *PLoS One, 9,* 1-5. doi: 10.1371/journal.pone.0109950

Appendix A: Comparative Conjunction Probability Judgement Task

	Description	Statement	Statement Type
Example 1	Henry is 44 years old. He lives in Africa and he is passionate about wild animals. He drives a jeep. He loves documentaries and adventure books. During his free time, he helps the inhabitants of a neighboring village build a well.	Henry is more likely to be a photographer and work in an animal testing lab than he is to be a photographer.	U&I
Example 2	Jesse is 28 years old. She enjoys meeting new people and traveling to unusual countries. She has excellent team building skills and is very athletic. In her spare time she likes to go hiking and play paintball.	Jesse is more likely to be a soldier than a beauty salon owner and a soldier.	R&L
Example 3	Julia is 35 years old. She plays kick-boxing. She is a member of the neighbourhood watch. She drives a truck. She has excellent problem-solving skills. She is confident and copes well with difficult situations.	Julia is more likely to be a florist and a firefighter than she is to be a florist.	R/I
Trial 1	Linda is 31 years old. She is single, outspoken, and very bright. She has a Degree in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.	Linda is more likely to be a feminist than she is to be a feminist and a cashier.	R&L
Trial 2	Johnny is 29 years old. He is a fan of punk and hard rock music. He likes to wear leather. He owns a Harley-Davidson and takes part in biker rallies every year. He has a Rottweiler named Rex.	Johnny is more likely to work as a bouncer and give ballet lessons than he is to give ballet lessons.	R/I

Trial 3	Ronald is 31 years old. He likes candlelight dinners and watching the sunset. He collects antique books and is particularly fond of medieval poetry. On Sundays, he enjoys taking his poodle for a walk in the countryside.	Ronald is more likely to be a war General than he is to write romantic novels and be a war general.	U/L
Trial 4	Sarah is 43 years old. Since her childhood, she has always been passionate about mystery stories. She has an intuitive mind and is very sensitive. She wears colourful clothes and many jewels. She likes to read romance novels.	Sarah is more likely to work as a foreman and be a fortune-teller than she is to be a fortune teller.	U&I
Trial 5	Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.	Bill is more likely to work as an accountant than he is to play in a rock band during his free time and work as an accountant.	R&L
Trial 6	Thomas is 10 years old. He gets along very well with his brothers and sisters. At school, he is very shy. His favourite pastimes are reading, drawing and solving puzzles. He dreams that he will become a doctor later on in life.	Thomas is more likely to be unruly and studious than he is to be unruly.	R/I
Trial 7	Christine is 25 years old. She likes to be alone and has few friends. She wears black, listens to metal music, and is fascinated by death. She likes to read vampire stories.	Christine is more likely to be a ballerina than she is to be a Goth and a ballerina.	U/L
Trial 8	Mike is 23 years old. He studied economics for his A-levels. He is bilingual	Mike is more likely to give chemistry lessons and work	U&I

	and likes to travel. He follows the news closely and is attuned to the latest happenings around him.	as a journalist than he is to work as a journalist.	
Trial 9	Chris is 19 years old. He has long hair and several piercings on his face. He has a tattoo on his right forearm. He participated in protests against genetically-modified crops. He works as a volunteer for a charity helping the homeless during winter.	Chris is more likely to be an anti-globalization activist than he is to work as a salesman and be an anti-globalization activist.	R&L
Trial 10	Susan is 35 years old. She is single, shy and rather unassuming. She owns her flat, which she has redecorated herself. She likes sewing, and is taking gardening classes. She owns four cats and volunteers for the RSPCA every Saturday.	Susan is more likely to be a librarian and enjoy bungee jumping than she is to enjoy bungee jumping.	R/I
Trial 11	Simon is 32 years. He works hard, loves money and speculates in stocks. He is often stressed, rarely has time to eat a full meal and sleeps little. He is single and lives in a loft apartment.	Simon is more likely to be a graffiti artist than he is to be a businessman and a graffiti artist.	U/L
Trial 12	Rachel is 25 years old. She is engaged. She wears the most fashionable, most expensive designer clothes. She is passionate about everything that relates to interior decorating. She works at a publishing house in Paris city centre.	Rachel is more likely to attend fashion shows and enjoy camping in the wilderness for her holidays than she is to attend fashion shows.	U&I
Trial 13	Karen is 40 years old. She is passionate about the environment and recycles all her waste. She lives a healthy lifestyle and is a	Karen is more likely to be an ecologist than she is to be an ecologist and enjoy taxidermy.	R&L

	vegetarian. She enjoys hiking and being surrounded by nature, and she is an active member of Green Peace.		
Trial 14	Samantha is 24 years old. She is a caring person who is always polite to others. She likes to help people and is very good at multi-tasking. She is clean, hygienic and remains calm in a crisis. She is energetic and has a great work ethos.	Samantha is more likely to be a nurse and a robotics engineer than she is to be a robotics engineer.	R/I
Trial 15	Nathan is 47 years old. He is lazy and slightly overweight. He enjoys DIY and is handy around the house. He likes to have a pint with the lads after work.	Nathan is more likely to be a personal trainer than he is to be a builder and a personal trainer.	U/L
Trial 16	Peter is 32 years old. He is a sports enthusiast who often travels with his friends to go and watch live sporting events. He likes to keep fit by jogging and has completed several long-distance races. He coaches little league football on the weekends.	Peter is more likely to be a rugby referee and interested in philosophy than he is to be a rugby referee.	U&I

Appendix B: Rational Experiential Inventory (Pacini & Epstein, 1999)

Rational-Experiential Inventory

Please use the following scale (1-5) to answer these questions:

1 = Definitely not true of myself; 5 = Definitely true of myself

1. I have a logical mind.
2. I prefer complex problems to simple problems.
3. I believe in trusting my hunches.
4. I am not a very analytical thinker.
5. I trust my initial feelings about people.

6. I try to avoid situations that require thinking in depth about something.
7. I like to rely on my intuitive impressions.
8. I don't reason well under pressure.
9. I don't like situations in which I have to rely on intuition.
10. Thinking hard and for a long time about something gives me little satisfaction.

11. Intuition can be a very useful way to solve problems.
12. I would not want to depend on anyone who described himself or herself as intuitive.
13. I am much better at figuring things out logically than most people.
14. I usually have clear, explainable reasons for my decisions.
15. I don't think it is a good idea to rely on one's intuition for important decisions.

16. Thinking is not my idea of an enjoyable activity.
17. I have no problem thinking things through carefully.
18. When it comes to trusting people, I can usually rely on my gut feelings.
19. I can usually feel when a person is right or wrong, even if I can't explain how I know.
20. Learning new ways to think would be very appealing to me.

21. I hardly ever go wrong when I listen to my deepest gut feelings to find an answer.
22. I think it is foolish to make important decisions based on feelings.
23. I tend to use my heart as a guide for my actions.
24. I often go by my instincts when deciding on a course of action.
25. I'm not that good at figuring out complicated problems.

26. I enjoy intellectual challenges.
27. Reasoning things out carefully is not one of my strong points.
28. I enjoy thinking in abstract terms.
29. I generally don't depend on my feelings to help me make decisions.
30. Using logic usually works well for me in figuring out problems in my life.

31. I think there are times when one should rely on one's intuition.
32. I don't like to have to do a lot of thinking.
33. Knowing the answer without having to understand the reasoning behind it is good enough for me.
34. Using my gut feelings usually works well for me in figuring out problems in my life.
35. I don't have a very good sense of intuition.

36. If I were to rely on my gut feelings, I would often make mistakes.
37. I suspect my hunches are inaccurate as often as they are accurate.
38. My snap judgements are probably not as good as most people's.
39. I am not very good at solving problems that require careful logical analysis.
40. I enjoy solving problems that require hard thinking.

Data scheme
Recode: 4, 6, 8, 9, 10, 12, 15, 16, 22, 25, 27, 29, 32, 33, 35, 36, 37, 38, 39

Rational Ability:=(1 + 4 + 8 + 13 + 14 + 17 + 25 + 27 + 30 + 39)/10
Rational Engagement:=(2 + 6 + 10 + 16 + 20 + 26 + 28 + 32 + 33 + 40)/10
Experiential Ability:=(3 + 5 + 18 + 19 + 21 + 34 + 35 + 36 + 37 + 38)/10
Experiential Engagement:=(7 + 9 + 11 + 12 + 15 + 22 + 23 + 24 + 29 + 31)/10

Appendix C: Experiment 1 Robustness test

Participants who failed to meet certain outlined parameters were flagged as outliers. This entailed examining the judgement responses, the judgement latencies and the reading durations for carelessness. Outliers were flagged for a number of reasons, which will be explained in more detail below, and following the number of flags received they were either retained in the sample, or excluded.

Following this, the total time taken to complete the task was screened for outliers in the unlimited time condition. A *z*-score was computed for each participant for the total time taken to complete all 16 trials in the study. Participants who performed slower or faster than 3 standard deviations from the mean total time were considered outliers. One participant in the unlimited time condition was identified as an outlier and flagged.

Next, the trial duration latencies were screened for outliers. In the limited time condition, participants were instructed to make their judgements in 8 seconds or faster. An average trial duration score was computed for each participant in this condition, and they received a flag if this was greater than 8 seconds. No participant spent over 8 seconds per trial under time pressure. In the unlimited time condition a *z*-score was computed for each average trial duration score. Participants who performed slower or faster than 3 standard deviations from the mean total time were considered outliers. One participant in the unlimited time condition was identified as an outlier and flagged.

Following this, the trial durations were analysed for multivariate outliers. Unlimited and limited time conditions were analysed separately. An examination of the Mahalanobis distance scores flagged two participants as outliers in the unlimited time condition and another two in the limited time condition. The same analysis was performed to identify multivariate outliers on trial reading durations (i.e., the length of time taken to read the scenario). However, this time the entire sample was analysed together because the time pressure manipulation did not affect the time allocated for reading the scenarios. The Mahalanobis distance scores indicated a total of nine participants flagged for giving answers considered as multivariate outliers on the reading durations.

Finally, an overall flag score was computed for careless responding. This entailed summing the flags accumulated across (a) incongruent answers to non-conflict trials, (b) taking too long or not long enough to complete the study, (c) taking too long or not long enough to read the scenarios, (d) taking too long or not long enough to read the statement and make a judgement, and (e) responding to the trials in under 8 seconds in the limited time condition. In the limited time condition, five participants each scored one flag, while one participant scored two flags. In the unlimited time condition, four participants each scored one flag, while two participants scored two flags. Participants who received two flags or more were removed from the sample. Thus, three participants in total were removed from the sample leaving the total amount of participants at 98 ($n_{unlimited}$ = 48; $n_{limited}$ = 50).

When heuristic scores were analysed using a 2 (time condition) x 2 (conflict condition) x 4 (trial blocks) mixed analysis of variance (ANOVA) with repeated measures on the last two factors, results were no different to those outlined above in chapter 3 (see Table 1).

Table 1

Mixed Analysis of Variance Summary Table for the Effects of Time Pressure and Block on Conflict Sensitivity

Source	*df*	*SS*	*MS*	*F*	*p*	Partial Eta Squared
Between subjects						
Time pressure	1	28,482.28	28,482.28	19.23	<.001	.167
Error 1	96	142,180.99	1,481.05			
Within subjects						
Block	3	914.22	304.74	0.52	.669	.005
Block x Time Pressure	3	2,202.48	734.16	1.25	.291	.013
Error 2	288	168,715.89	585.82			
Conflict Sensitivity	1	12,751.92	12,751.92	13.13	<.001	.120
Conflict Sensitivity x Time Pressure	1	4,295.29	4,295.29	4.42	.038	.044
Error 3	96	93,268.49	971.55			
Block x Conflict Sensitivity	3	1,696.02	565.34	1.33	.264	.014
Block x Conflict Sensitivity x Time Pressure	3	458.77	152.93	0.36	.782	.004
Error 4	288	122,232.55	424.42			

Judgement latencies were converted to $\log^{10}$ prior to analysis to normalize the distribution and subjected to a 3(judgment type: no-conflict vs. conflict-heuristic vs. conflict-logical) x 2(time condition: unlimited vs. limited) mixed ANOVA with repeated measures on the first factor. Results do not differ from those reported above which include the full sample (see Table 2).

Table 2

Mixed Analysis of Variance Summary Table for the Effects of Time Pressure on Response Type

Source	*df*	*SS*	*MS*	*F*	*p*	Partial Eta Squared
Between subjects						
Time pressure	1	14.26	14.26	171.16	<.001	.74
Error 1	61	5.08	0.08			
Within subjects						
Response Type	2	0.18	0.09	8.28	<.001	.12
Response Type x Time Pressure	2	0.12	0.06	5.51	.005	.08
Error 2	122	1.31	0.01			

Appendix D: Barratt Impulsiveness Scale Version 11

DIRECTIONS: People differ in the ways they act and think in different situations. This is a test to measure some of the ways in which you act and think. Read each statement and put an X on the appropriate circle on the right side of this page. Do not spend too much time on any statement. Answer quickly and honestly.

①	②	③	④
Rarely/Never	Occasionally	Often	Almost Always/Always

1	I plan tasks carefully.	①	②	③	④
2	I do things without thinking.	①	②	③	④
3	I make-up my mind quickly.	①	②	③	④
4	I am happy-go-lucky.	①	②	③	④
5	I don't "pay attention."	①	②	③	④
6	I have "racing" thoughts.	①	②	③	④
7	I plan trips well ahead of time.	①	②	③	④
8	I am self controlled.	①	②	③	④
9	I concentrate easily.	①	②	③	④
10	I save regularly.	①	②	③	④
11	I "squirm" at plays or lectures.	①	②	③	④
12	I am a careful thinker.	①	②	③	④
13	I plan for job security.	①	②	③	④
14	I say things without thinking.	①	②	③	④
15	I like to think about complex problems.	①	②	③	④
16	I change jobs.	①	②	③	④
17	I act "on impulse."	①	②	③	④
18	I get easily bored when solving thought problems.	①	②	③	④
19	I act on the spur of the moment.	①	②	③	④
20	I am a steady thinker.	①	②	③	④
21	I change residences.	①	②	③	④
22	I buy things on impulse.	①	②	③	④
23	I can only think about one thing at a time.	①	②	③	④
24	I change hobbies.	①	②	③	④
25	I spend or charge more than I earn.	①	②	③	④
26	I often have extraneous thoughts when thinking.	①	②	③	④
27	I am more interested in the present than the future.	①	②	③	④
28	I am restless at the theater or lectures.	①	②	③	④
29	I like puzzles.	①	②	③	④
30	I am future oriented.	①	②	③	④

Patton, Stanford, Barratt (1995). *J Clin Psy*, vol. 51, pp. 768-774

Reverse scored items: 1, 7, 8, 9, 10, 12, 13, 15, 20, 29, 30.

Appendix E: Cognitive Reflection Task (CRT; Frederick, 2005)

1. A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents

   [Correct answer 5 cents; intuitive answer 10 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

   [Correct answer 5 minutes; intuitive answer 100 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 of the lake? ____ days

   [Correct answer 47 days; intuitive answer 24 days]