The Bright Homunculus in our Head: Individual Differences in Intuitive Sensitivity to Logical Validity

Omidreza Ghasemi

Student ID: 45050007

Department of Cognitive Science, Macquarie University

Supervisors:

Professor Simon Handley

Dr. Stephanie Howarth

Empirical thesis submitted in partial fulfilment of the requirements for the degree of Master of Research (Human Sciences),

Date submitted: 20th April, 2019

Abstract

Classical dual process theories of human reasoning attribute explicit reasoning to effortful, deliberative thinking. Lacking any access to the formal rules of logic and probability, according to these models, intuitive processes rely exclusively on superficial features of the problem. In recent years, however, researchers have demonstrated that reasoners are able to solve simple logical or probabilistic problems relatively automatically, a capability which has been called logical intuition. In two experiments, we examined the existence of this capability by instructing participants to rate their judgment of likeability (Experiment 1) and brightness (Experiment 2) of several reasoning problems. In order to investigate individual differences in these measures of logical intuition, participants were also asked to complete two scales of cognitive ability and cognitive style. The results showed that participants rated the conclusion to logically valid statements more likable and more physically bright. Although participants with higher cognitive ability showed greater logical intuition in their liking judgment, this effect, however, was absent when the brightness task, as another measure of logical intuition, was used. We discuss the implications of our findings for recent dual process theories of human reasoning.

Keywords: dual process theory, logical intuition, liking ratings, brightness, cognitive ability, analytic cognitive style

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would first like to thank to my supervisors, Professor Simon Handley and Dr. Stephanie Howarth, with whom I have had the pleasure to work on this project. I am grateful for all of their effort in helping me, not only in conducting this experiment, but also in learning how to think critically.

I would also like to thank to Dr. Dries Trippas and the PsychoPy team, especially David Bridges and Michael MacAskill for all of their generous help in designing the experiment.

Thanks to my friends, Hamid, Luke, Phillip, and Ehsan, for all of their supports during these months.

Last but not least, thanks to my mother, Mahboobeh, for the encouragements that made me to follow my dreams.

Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Omidreza Ghasemi

Chapter 1: Literature Review

Introduction

In 1983, Tversky and Kahneman asked several questions from different groups of participants. One of the questions was the well-known Linda problem:

> *"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 also participated in anti-nuclear demonstrations."* (Tversky & Kahneman, 1983)

Participants were asked to rank several social classes based on the resemblance of Linda to a typical member of those classes. Among the classes were "*Linda is a bank teller"* and *"Linda is a bank teller and is active in the feminist movement."* Regardless of their knowledge of and education in probability, more than 85% of the participants chose the latter (Linda is a bank teller *and* active in the feminist movement) as more representative of Linda's description. According to the authors, participants mainly engaged in intuitive thinking and judged based on some superficial heuristics, rather than thinking deliberatively about probability rules, which results in the ignoring of the normatively correct option (Linda is a bank teller). This distinction between intuitive thinking and deliberative thinking and their interaction is illustrated in a famous description by Stephen Jay Gould (1989):

> *"I know that the third statement…* [bank teller *and* active in the feminist movement] *is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me - 'but she can't just be a bank teller; read the description"."*

That there exists a distinction between intuition and reasoning is a dominant view amongst lay people and within scientific circles. The traces of this distinction can be dated back to the early philosophical accounts of human mind, even to Plato's writings, and early conceptualizations of the human psychology, such as Freud's and James' theories (Frankish & Evans, 2012). This view considers intuition as a mechanism which depends on feelings and is void of any logical consideration. Thus, it is an unreliable source of thinking which should be overruled by logical thought in order to reach optimal performance in reasoning, judgment, and decision-making and to avoid the well-known cognitive biases.

The idea of two distinct processes in one mind continues to dominate recent scientific theories of human reasoning. An example of these theories is the generic dual process theory, and more specifically, the dual process theory of reasoning (Evans, 2008; Sloman, 1996; Stanovich & West, 1998). According to dual process accounts, two different types of processes are involved in judgment and reasoning: Type 1 (T1) or intuitive processes and Type 2 (T2) or deliberative processes (Evans, 2003). Although several theorists have rejected attributing biased thinking to intuition and also the normative responding to deliberation (Evans, 2017; Evans & Stanovich, 2013; Pennycook, De Neys, Evans, Stanovich, & Thompson, 2018, see also, Gigerenzer & Gaissmaier, 2010), the failure to engage in deliberative thinking has been assumed as the main factor explaining incorrect responses (Epstein, 1994; Kahneman & Frederick, 2005; Morewedge & Kahneman, 2010).

In the following review, we discuss the distinction between intuitive and deliberative reasoning in more depth. Specifically, we are interested in investigating the characteristics of each type to find out whether there is a necessary association between types of thinking and response accuracy in logical reasoning. In other words, can we attribute the ability to reason logically to our intuitive thinking or, as the common distinction of reason and intuition implies, is it an exclusive feature of deliberative processes? Several findings have revealed that people are able to solve reasoning problems intuitively, a capability that has been called the "logical intuition" or intuitive logic (De Neys, 2012; Thompson & Newman, 2017). More importantly, if people can detect the underlying logical structure of a problem intuitively, is this capability different among people with higher and lower cognitive abilities? The second question is related to a more general query that investigates the factors that differentiate good and bad reasoners - the divergence problem in reasoning and decision-making (De Neys & Bonnefon, 2013). At what points in the reasoning process does the divergence between reasoners happens? If people with high and low cognitive ability (CA) have a similar intuitive logic capacity, it indicates that good and bad reasoners diverge at the late stages of the reasoning process and their deliberative thinking determines their performance on such problems. On the other hand, if reasoners showed individual differences in their intuitive logic, it reveals that divergence takes place at the beginning of the reasoning process. Thus, are there any differences among participants with different CA and analytic cognitive style (ACS)[1] regarding their logical intuition?

To address these questions, we review the main theoretical models of dual process theories and their position in each of the questions above. We start from *the default interventionist (DI) model* and its different versions. Then, we discuss several accounts that have been developed to explain recent incompatible findings with the DI account. These models, which have been collectively referred to as the dual process theories 2.0 (De Neys, 2017), include *the logical intuition model*, *the three-stage dual process model*, and *the parallel processing model*, among others. In explaining each model, we consider the possibility of intuitive logic, whether reasoners have an intuitive sensitivity to logic, and the existence of individual differences regarding this capability. The final section in this chapter provides a summary of the aims and hypotheses of the current study.

The Default Interventionist model

One of the most influential theories among dual process theories of reasoning is the DI model. Consistent with the generic dual process theories, the DI model assumes that our reasoning apparatus has been comprised of two qualitatively distinct processes, namely, intuitive T1 processes and deliberative T2 processes (Evans, 2008). To define the characteristics of each process, recent theories place an emphasis on the defining features which are necessary criteria to distinguish between these types of thinking. Several studies have also revealed some incidental correlates of each type; however, these features do not necessarily separate these two types of thinking. The defining feature of T1 processes is automaticity, and their correlated features are parallel, unconscious, and associative processing as well as the independence from cognitive capacity. The defining features of T2 processes are dependence on working memory, and the ability to decouple and their typical correlates are controllability, consciousness, and dependence on cognitive capacity (Evans & Stanovich, 2013; Pennycook, De Neys, Evans, Stanovich, & Thompson, 2018, but see, Melnikoff & Bargh, 2018).

[1] Cognitive ability is the ability to use deliberative thinking when faced with situations that demand such thinking, while analytic cognitive style is the willingness and inclination to engage in deliberative thinking instead of a fast and less-demanding intuitive thinking. We will discuss about these two measures in more details in the following sections.

Based on the DI model, due to the effortless and automatic nature of the intuitive processes, T1 processes precede T2 processes and provide an initial response. At the later stage, depending on the characteristics of the stimuli and top-down cognitive and motivational factors, reasoners may engage in the more effortful T2 processes. These deliberative processes may either confirm, modify, or override the initial responses of T1 processes (Kahneman & Frederick, 2002), or as Evans (2017) put forward, evaluate the justifiability of the initial mental representation. The precedence of T1 processes, or the speed asymmetry assumption, is a core concept in the DI account that, despite the different revision of this theory in recent years, still exists in the latest DI account (Evans, 2018).

The speed asymmetry assumption has been used commonly to account for different phenomena that have emerged in recent studies. One of the most studied of these kinds of phenomena is belief bias, which refers to the tendency to evaluate an argument based on its believability instead of its logical validity, even in the presence of strong instructional emphasis on the logical necessity rule (Evans, Barston, & Pollard, 1983; Evans, Handley, & Harper, 2001; Newstead, Pollard, Evans, & Allen, 1992). For example, consider the following syllogistic reasoning problem from De Neys and Franssens (2009):

"All flowers need water
Roses need water
Roses are flowers"

Based on our knowledge and experience from the real world, all flowers, including roses, need water. So, judging based on our prior knowledge, reasoners show a tendency to base their judgment on their belief and endorse this problem as logically valid regardless of the fact that the conclusion of this reasoning problem does not logically follow from its premises. The structure of this problem is relatively straightforward. Nevertheless, reasoners, especially those with lower CA, seem to have difficulty ignoring their belief-based processes in favor of logical rules on problems of this kind (Stupple, Ball, Evans, & Kamal-Smith, 2011).

One of the implications of the speed asymmetry assumption is the necessary correspondence between responses (belief-based or logic-based) and underlying processes (T1 or T2). By assuming a correspondence between the responses and underlying processes, the DI account attributed belief-based judgment to automatic T1 processes and logic-based judgment to analytic T2 processes (Evans & Stanovich, 2013). Belief-based judgment occurs rapidly and automatically, and logic-based judgment demands cognitive decoupling from the context of knowledge. Thus, logic-based judgment, at least on the belief bias task and some similar paradigms, is mainly exclusive to T2 processes. On the other hand, T1 processes rely primarily on superficial cues such as the matching of different elements of the problem (e.g., matching bias), stereotypical characteristics (e.g., base-rate neglect), believability and acceptability of the content (e.g., belief bias and myside bias), and co-occurrence information and associative representations (e.g., conjunction fallacy; Klauer, Musch, & Naumer, 2000; Sloman, 1996; Tversky & Kahneman, 1983).

Numerous studies provide evidence supporting these direct correspondences and predictions of the DI model. For example, by minimizing the engagement of T2 processes by various experimental manipulations such as limiting the response time (Evans & Curtis-Holmes, 2005) and loading working memory capacity (De Neys, 2006), researchers revealed an increase in belief-based responses. Moreover, belief-based responses are associated with higher activation of the ventral medial prefrontal cortex (VMPFC), a region which has been shown to have a role in emotional processing, while logic-based responses have a relationship with higher activation of the right lateral prefrontal cortex (RLPFC), a region responsible for cognitive monitoring (Goel & Dolan, 2003). Finally, Individual differences studies (see, for example, Stanovich & West, 1998) support the DI model by showing that reasoners with the higher CA and analytical thinking disposition[2], typical markers of successful T2 thinking, showed lower levels of belief bias. Similar to reasoners with lower CA, younger children have a greater tendency to base their judgment on belief than the logical structure of the problem (Kokis, Macpherson, Toplak, West, & Stanovich, 2002).

In contrast to the initial conceptualizations, in the later revisions, the DI model theorists have attempted to account for claimed T1 sensitivity to several logical rules. For example, Evans (2017, 2018) suggested that reasoners are able to solve some simple reasoning problems through T1 processes. Stanovich (2018) suggested a mindware continuum hypothesis with three parameters of mindware (storage), detection and override, according to which, people have a varying level of knowledge across different reasoning problems. At one end of the continuum (i.e., the mindware gap area), we lack the appropriate knowledge regarding a problem, which at the later stage, leads to the failure of the detection and the override of an intuitive response. On the other hand, some rules and structures are practiced during development and become automatic. These automatic logical rules are inherently intuitive and are able to cue normative responses without the engagement of T2 processes. In other words,

[2] Cognitive ability is the "ability" to engage in deliberative thinking, while analytic thinking disposition (or cognitive style) is related to the willingness to doing so.

the automatized knowledge facilitates the detection and override of non-normative responses in the level of T1 processes. This automatic knowledge storage is an indication of intuitive logic that has been proposed by the recent dual-process theories (Bago & De Neys, 2017a; Morsanyi & Handley, 2012a; Thompson & Newman, 2017; Trippas, Handley, Verde, & Morsanyi, 2016)

Similar to the DI model, the parallel competitive model, as another classic dual process theory of reasoning, distinguishes between an associative-based system and a rule-based system. According to Sloman (1996), activated simultaneously at the beginning of the reasoning process, the intuitive system relies primarily on similarity and co-occurrence information, whereas deliberative thinking can represent the rules and structures, especially causal structures. However, in a significant revision of this model, Sloman (2014) retracted this assumption, amongst others, and attributed some levels of causal structure representation to intuitive processes. For example, Hagmayer and Sloman (2009) showed that people rarely endorse the choice of buying running shoes to lose weight from the information that there is a relationship between buying running shoes and getting into better shape, and they reject this choice relatively automatically. In conclusion, despite the correspondence between type of response and underlying thinking in many dual process models (e.g., Dhar & Gorlin, 2013; Epstein, 1994; Kahneman & Frederick, 2002), other DI theorist incorporate the automatic and intuitive logic into their model of reasoning (e.g., Evans, 2017, 2018; Sloman, 2014; Stanovich, 2018).

Regarding the individual differences in T1 and T2 thinking, the DI models generally assume an asymmetrical correlation between the type of processes and CA and ACS; T2 deliberative thinking depends on CA whereas T1 processes are independent of CA, working memory and executive functions (Evans & Stanovich, 2013; Stanovich, 1999; Stanovich & West, 2008). As Evans (2008) described, it seems that two separate systems control the behavior, the deliberative system with the dependence on executive functions and the intuitive one with independence from these resources. In a similar vein, Stanovich and West (2008) showed that CA is independent of the performance in most of the tasks designed to examine biased thinking. Based on these findings, these authors developed a 3-parameter model; mindware (storage), detection and override. According to this model, CA is only able to differentiate reasoners in the override stage and it is not a good predictor of performance in the first two parameters. The required knowledge (i.e., mindware) to tackle reasoning problems is acquired through the past learning and experience for reasoners with both high and low CA. Thus, the differences in performance in the reasoning process are mainly raised from the decoupling ability of reasoners with higher CA. Similarly, the parallel competitive model considers deliberative thinking as the only source of individual differences (Darlow & Sloman, 2010). However, in his recent writing, Evans (2017) proposed that individual differences can also affect T1 thinking, as studies have shown that people with higher CA have different response patterns in the tasks aimed to evaluate T1 thinking (For example, see, Newstead, Handley, Harley, Wright, & Farrelly, 2004; Thompson, 2014; Thompson, Pennycook, Trippas, & Evans, 2018).

Logical Intuition

In a series of studies, De Neys and his colleagues presented participants with conflict and no-conflict versions of different reasoning tasks. A conflict problem is a kind of problem in which the intuitive response is at odds with the response cued by deliberative processes. An example of such a problem is similar to the syllogistic problem that was presented in previous pages. On that problem, a response based on the T1 processes (belief-based response) is inconsistent with a response based on T2 processes (logic-based response). A no-conflict version of that problem is one in which its validity and believability cue a similar response. For example, the following problem is a no-conflict valid and believable version of the previous conflict syllogism used in conflict detection studies (De Neys, 2014):

"All flowers need water
Roses are flowers
Roses need water"

The primary purpose of these studies was to answer the following question: Do people, especially when they give biased responses, detect the conflict between their incorrect responses and a normative response? According to De Neys and his colleagues, when trying to answer conflict problems compared to no-conflict problems, participants show an increase in response time (De Neys & Glumicic, 2008; Frey, Johnson, & De Neys, 2017), increase in skin conductance responses (De Neys, Moyens, & Vansteenwegen, 2010), an increased tendency to re-view critical elements of the problem (De Neys & Glumicic, 2008), a decrease in the confidence level (De Neys, Cromheeke, & Osman, 2011; Thompson & Johnson, 2014), and an increase in the activation of anterior cingulate cortex (De Neys, Vartanian, & Goel, 2008) and centro-parietal N2 and frontal P3 (Bence Bago et al., 2018), regions which have been attributed mainly to the conflict detection sensitivity. These findings were consistent across different problem types (Ball, Phillips, Wade, & Quayle, 2006; Bonner & Newell, 2010; De Neys & Franssens, 2009; De Neys, Rossi, & Houdé, 2013; Morsanyi & Handley, 2012; but see, Pennycook, Fugelsang, & Koehler, 2012; Travers, Rolison, & Feeney, 2016).

Where is this conflict sensitivity coming from? Is this a conflict between an intuitive T1 process and a deliberative T2 process? Recent evidence and theoretical conceptualizations emphasize that the conflict is between two alternative T1 responses. If the conflict is between T1 and T2 responses, the question arises as to what causes the initiation of T2 processes in the first place, since it cannot be produced automatically by the stimulus (Pennycook, 2017). The need of T2 processes for a trigger or an initial cause indicates that the conflict is between, at least, two T1 responses. This assumption is supported by evidence from implicit measures (such as skin conductance measures) in previous studies (De Neys et al., 2010). Detecting the conflict on an implicit level of inference indicates that reasoners can automatically evaluate the consistency of their initial responses with the correct normative responses without engaging in T2 processes. These findings strengthen the claims that sensitivity to logical validity is partly implicit and thus is related to intuitive rather than deliberative processes (Bence Bago & De Neys, 2017b; De Neys, 2014).

A body of experimental findings has shown that T1 thinking can produce the correct normative answer without any engagement of T2 processes. In fact, it is not necessary for reasoners to delay their responses until the involvement of the corrective T2 processes because they often reach the correct answer by using their fast T1 processes. For example, in 7 studies, Bago and De Neys (2018) demonstrated that participants who answered reasoning problems correctly, those answers were generally raised from intuitive T1 thinking, rather than a corrective T2 deliberative system. Similarly, Bago and colleagues (2019a) found that when biased participants were asked to provide a second guess besides their initial answer to a mathematical reasoning problem, their second guess was closer to the correct answer.

To explain findings of this kind, De Neys (2012) proposed *the logical intuition model*, according to which a problem cues two T1 responses: an intuitive heuristic response and an intuitive logical one. Depending on the reasoner's underlying knowledge of the rule of probability and logic (Frey et al., 2017) and different problem features such as the saliency of critical information (Bago & De Neys, 2017a), the activation of both responses on conflict problems results in a conflict detection, which in turn, increases the likelihood of engaging in T2 processes. Thus, the intuitive sensitivity to logical validity is a core principle in the logical intuition model. This sensitivity reflects the successful conflict detection between an intuitive heuristic response and an intuitive logical response. To test these assumptions, Bago and De Neys (2019b) manipulated the strength of the logical intuition by reducing strength of the problem features that cue such an intuition (e.g., the base-rate information) and found that this manipulation reduced the likelihood of normative responding in the absence of T2 thinking. There is also a boundary condition, however, that we need to consider when discussing conflict detection. Conflict detection has been demonstrated in simple reasoning problems, and this effect tends to be weak or absent in more difficult problems (Brisson, Schaeken, Markovits, & De Neys, 2018).

To address the individual differences debate, one can ask whether reasoners have different conflict detection abilities, and hence, different logical intuition capacity. In order to answer this question, researchers have distinguished between three stages of the reasoning process which were described earlier: storage, detection and override. According to the early accounts of the logical intuition model, since conflict detection is highly efficient and even the most biased reasoners can detect the conflict, biases are the consequences of override failures rather than lax monitoring (De Neys, 2017; De Neys & Bonnefon, 2013). The efficient conflict detection and the inefficient override imply that what determines the accuracy of the final response does not appear at the first stage of the reasoning. In other words, the individual differences influence reasoners at the later stage of reasoning (inhibition) rather than in the early stages (storage or conflict detection; De Neys, 2015).

This assumption is contradicted by a study of Pennycook, Fugelsang, and Koehler (2015) who tested *the three-stage model of dual process theory*. According to this model, at the beginning of the reasoning process, varying features of a stimulus cue several initial responses (Stage 1) and their possible conflict may or may not be detected (Stage 2). Finally, the final answer will be selected, either by engaging in T2 thinking or sticking to the initial intuitive response (Stage 3). In 4 experiments, Pennycook et al. (2015) showed, not only that the efficient conflict detection assumption is indefensible, but also that reasoners differ in their sensitivity to the conflict signals. The most biased subset of the participants failed to detect the conflict between two opposing responses (see Experiment 1 and 2). Moreover, similar to the previous studies (e.g., Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014a), reasoners with a higher score in analytic thinking style scales showed greater detection of the conflict. These findings imply that individual differences could influence both T1 and T2 processes.

Following these findings and other evidence of early divergence (Mevel et al., 2015), a recent revision of the logical intuition model considers the possibility of conflict detection failure. According to Frey and colleagues (2017), while most of the reasoners were able to detect the conflict successfully, there was a subgroup of biased reasoners (around 12%) who failed to recognize the conflict as reflected by their response time and confidence ratings across a variety of reasoning problems. It seems plausible that, at least for a subgroup of reasoners, individual differences influence the reasoning process relatively early. The authors also showed that one cognitive factor that can affect the likelihood of conflict detection is possession of relevant knowledge (i.e., storage).

The Parallel Processing Model

Despite a large body of supportive evidence (For a critical review, see, Ball, Thompson, & Stupple, 2017), there are several findings which are inconsistent with the predictions of the DI model. Regarding the attribution of belief-based responses to T1 thinking and logic-based responding to T2 thinking in the belief-bias paradigm, recent evidence has revealed that such a direct correspondence is unsound. For example, time pressure manipulations to prevent the engagement of T2 thinking does not affect the rate of belief-based responses in conditionals (Evans, Handley, & Bacon, 2009), no necessary link has been found between lower cognitive capacity and higher belief-based responses (Klaczynski, 2000; Newstead et al., 2004), and belief bias was more common among children with higher age compared to younger children, especially in children with higher CA (Morsanyi & Handley, 2008). Considering all these findings, it seems plausible that belief-based judgments are not solely dependent on T1 processes.

On the other hand, several studies in different fields showed that logic-based judgment could be accomplished via T1 thinking. For example, text comprehension studies have demonstrated that reasoners draw a conclusion from premises even when this inference is not necessary for the comprehension of the text coherency (Lea, O'Brien, Fisch, Noveck, & Braine, 1990; Leo & Greene, 2008). Lea (1995) demonstrated that readers were slower at assessing a target word as an English word, when the target word (e.g., "BUTTER") was followed by a relevant inferential text ("Tony need to choose between bread and corn flake, and he decided not to have corn flake") compared to a relevant no-inferential text ("Tony need to choose between bread and corn flake, and he decided to talk to her mom about corn flake"). This finding, according to the author, supports the idea that readers automatically make propositional inferences when premises are available. Similarly, Rader and Sloutsky (2002) presented major and minor premises in the context of a story and showed that participants automatically and falsely endorsed that the conclusion had been presented in the story. Finally, Reverberi and colleagues (Reverberi, Pischedda, Burigo, & Cherubini, 2012) presented the second premise ("number 3") of a reasoning problem in an undetectable speed following by the first premise ("If 3 then 8"). Participants were relatively faster at evaluating a target number ("number 8") as odd or even when the second premise contained a number presented in the first premise, thus creating a valid problem, compared to the conditions in which the second premise contained a different number ("number 5"). These findings indicate that people can process the basic logical structures relatively fast and automatically, without engaging in the T2 deliberative thinking.

The aforementioned findings of effortful belief and automatic logic pose challenges to the DI account, in specific, and dual process theories, in general (Howarth & Handley, 2017). To account for these inconsistent findings, Handley and Trippas (2015) proposed *the parallel processing model*, according to which there is not a direct correspondence between belief and logic and processing type. Belief-based judgments are no longer assumed as sole products of T1 processes, thus, access to structural features of a problem such as its logical validity is not an exclusive characteristic of T2 processes.

To test the predictions of the parallel processing model, Handley and colleagues used a different instructional manipulation compared to those of the traditional reasoning studies. In the classic reasoning studies, participants are supposed to judge whether a conclusion necessarily follows from the premises (see, for example, Evans et al., 2009; Newstead et al., 1992). Handley et al. (2011) asked participants to judge conflict and no-conflict problems based on either the logical validity or believability of the presented conclusion. The DI model predicts that reasoners should have higher accuracy and lower latency in the belief instruction compared to the logic instruction. Contrary to this prediction, the findings revealed that reasoners under the logic instruction had greater accuracy and lower latency. More importantly, logic interfered with belief (the conflict between validity and believability of the problem had a pronounced effect under belief instruction), while belief did not interfere with logic to the same extent. These findings which replicated across five experiments, with within-subject and between-subject designs, suggest that, at least on some levels, logical inferences depend on automatic processes and belief evaluations depend on slower and more effortful processes.

In a similar vein, Pennycook, Trippas, Thompson, and Handley (2014) employed a similar instructional manipulation for base-rate problems. Reasoners were asked to decide between two social groups ("accountant and street artist") that a person (e.g., "Brannon") may belong to, based on the statistical information ("995 accountants and 5 street artists") or the personality description of that person (e.g., "being good with numbers"). Participants were instructed to respond based upon either the statistics or the description. The DI model of reasoning would predict that only the personality description should interfere with the statistics since a response based on this feature should be accomplished faster via T1 thinking. However, researchers showed that both personality description and statistical information interfere with each other to a similar degree. In another study, Howarth, Handley, and Walsh (2016) imposed a cognitive load on participants by asking them to generate random numbers while they are answering reasoning problems. Again, consistent with the PP model, cognitive load had a detrimental effect on both belief-based and logic-based judgments. Finally, Howarth, Handley, and Walsh (2018) increased the inhibitory demand of their tasks by asking participants to judge the believability, validity and some physical characteristics of the problem such as color and font style. They showed that this increased demand affected the belief judgment more than the logical judgment. The effect of cognitive load on both kinds of judgments indicates that these inferences do not necessarily correspond with distinct processes (Trippas & Handley, 2018).

The evidence of automatic/intuitive logic is not exclusive to studies aligned with the parallel processing account. For example, Newman, Gibb, and Thompson (2017) found that reasoners can make rule-based inferences even when they are required to answer in a limited time. They also incorporated belief-based inferences in their deliberative thinking when they were given an unlimited time to revise their initial intuitive answer. These findings indicate that reasoners can make inferences based on the logic and probability rules relatively fast (via T1 thinking) and make belief-based inferences relatively slowly (via T2 thinking).

According to the parallel processing model, the interference of each problem feature (believability and validity) depends on the speed with which one is completed. In simple reasoning problems (like modus ponens conditionals and disjunctions), a logic-based response would be generated early (consider the ease with which one can solve problems in the form of "If P then Q; P; Therefore Q"), and it is relatively immune to the interference of belief-based responses. On the other hand, in complex problems (such as multiple model syllogisms), a belief-based response is more accessible than a logic-based response (consider the relative difficulty with which one solves a problem in the form of "Some A are B; No B are C;

Therefore, some C are not A"). Consistent with this prediction, Trippas, Thompson, and Handley (2017) found that on simple problems (such as modus ponens), logic-based judgments were more accurate and logical structure interferes with belief judgments. This interference was bidirectional on moderately complex reasoning problems (such as modus tollens and single model syllogisms). Finally, on the most complex reasoning problems (multiple model syllogisms), believability interferes with logic.

Since the parallel processing model is still in its infancy, it does not directly address the individual differences debate. It has been long established that people with higher CA have a better performance in T2 processing (Toplak, West, & Stanovich, 2011, 2014). Since this model attributes some previously thought functions of T2 thinking to T1 processes, it is probable that some of the differences between high and low CA participants in reasoning problems arise from their differences in T1 thinking. Some support for this hypothesis comes from Thompson and colleagues' studies. For example, Thompson et al. (2018) used instructional manipulations along with different CA and ACS measures (i.e., Shipley-2 test, numeracy test, cognitive reflection test, and actively open-minded thinking questionnaire). The findings showed that, for participants with higher ability, logic-based responses were more accurate and more accessible than belief based-responses. This indicates that, for this group of reasoners, relying their judgment on the logical structures of the problem is the default process, even in the absence of T2 thinking. The pattern of the findings was reversed for participants with lower CA.

Similarly, Thompson and Johnson (2014) used a two-response paradigm in which participants were presented with a problem twice; once under limited time and once with no time pressure. The authors demonstrated that even when participants were asked to respond as quickly as possible, those with higher CA and ACS are doing better to respond based on the logical rules. According to these authors, individual differences between high and low ability reasoners arise early in the reasoning stage and they may have little differences regarding conflict detection and override.

These results, however, were contradicted by the experiments of Morsanyi and Handley (2012). They showed that the intuitive sensitivity to logical validity is independent of CA. They showed that validity judgment (i.e. judgment of the validity of a problem; a T2 demanding task) depends on the working memory capacity as measured by the operation span task and the figure of the reasoning problem, whilst the liking judgment (i.e., rating the likeability of a problem; an index of T1 thinking) was not affected by these factors (Experiment 2 and 3). Similarly, Nakamura and Kawaguchi (2016) did not find a significant correlation between thinking dispositions as measured by Rational-Experiential Inventory and liking judgments. This finding indicates that intuitive logic, like other T1 processes, is independent of cognitive capacity. On the other hand, these researchers demonstrated that intuitive liking judgments were related to explicit reasoning; reasoners with better performance in the explicit reasoning task had higher liking rating for valid problems. These contradictory findings make it hard to extract any straightforward principle from the parallel processing model regarding the relationship between cognitive abilities and individual differences in intuitive logic.

Coherency, Fluency and Intuitive judgment

One implication of the ability of T1 thinking to process the logical structure of a problem is that people have a sensitivity to the underlying logical structures when faced with an argument (Thompson & Newman, 2017). Different methods have been proposed to study intuitive logic such as conflict detection, the two-response paradigm, and the second-guess method, that we reviewed earlier. The fluency misattribution paradigm suggests another legitimate method to investigate logical intuition (Reber & Unkelbach, 2010; Topolinski & Reber, 2010; Winkielman, Schwarz, Fazendeiro, & Reber, 2003).The *fluency-affect intuition model* (Topolinski, 2011, 2018) assumes that coherency leads to higher levels of fluency, which in turn, due to hedonically marked experience, results in positive affect. This affect, at the later stages of the process, may be used as a cue for intuitive judgment (see, Figure 1). In a series of studies, researchers created semantic coherent word triads (e.g., FRESH, HOLY, and LIQUID with a common target: WATER) and incoherent word triads (e.g., DREAM, BALL, BOOK with no readily common target). Participants were able to distinguish coherent and incoherent triads above the chance level (Topolinski & Strack, 2009). This implicit coherency also influences liking rating (Reber & Schwarz, 1999), positive affect (Topolinski, Likowski, Weyers, & Strack, 2009), and judgments of truth (Topolinski & Reber, 2010).

By applying the fluency-affect paradigm to the reasoning domain, researchers have investigated intuitive sensitivity to logical validity. For example, Morsanyi and Handley (2012) asked participants to rate their liking, and in a different block, to judge the logical validity of several reasoning problems. In 4 experiments, they showed that participants liked and endorsed believable and valid problems more than unbelievable and invalid ones. Unlike endorsement ratings, liking ratings were not affected by instruction (Experiment 1), the figure of the syllogisms (Experiment 2), and participants' working memory capacity (Experiment 3). Moreover, misattribution manipulation (i.e., background music) and affective priming (Experiment 4) influenced the liking ratings but not the validity judgment. These findings imply that, according to Morsanyi and Handley, reasoners have an intuitive capacity to detect the logical validity of problems.

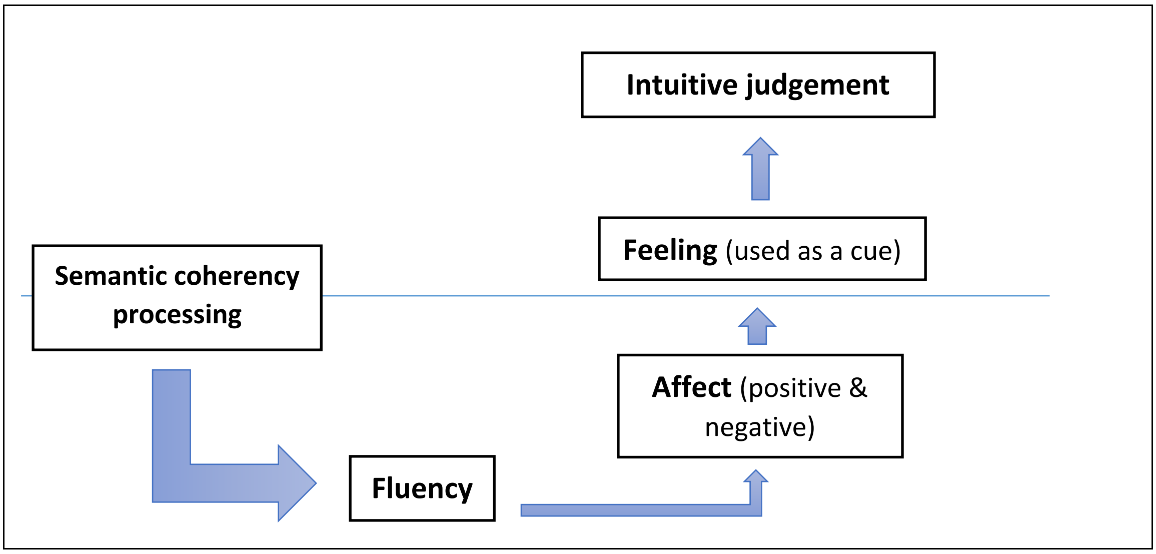

Figure 1 An illustration of the fluency-affect intuition model proposed by Topolinski (2011)

Similarly, Trippas, Handley, Verde, and Morsanyi (2016) presented participants with different problem types and asked them to rate their liking (Experiment 1) and evaluate the brightness of the conclusion to each problem (Experiments 2 & 3). They found that participants judged logically valid and believable statements more likable and brighter than invalid and unbelievable statements. Moreover, a simulated smile (putting a pen between the lip) affected the judgment of brightness (Experiment 3). Whilst the Morsanyi and Handley study has been criticized for failing to adequately control content across problem structures (Klauer & Singmann, 2013; Singmann, Klauer, & Kellen, 2014), Trippas et al. (2016) used a variety of reasoning problems (conditionals, disjunctions, and syllogisms), and an implicit task (brightness judgment) that has no direct connection to the logical judgment. Based on these findings, Trippas et al. (2016), concluded that the logical structure of the problem is readily accessible for intuitive processes (See also, Morsany & Handley, 2012).

The Present Study

The core aim of the present study is two-fold. First, it investigates the existence of intuitive sensitivity to logic across several experimental tasks, instructions, and reasoning problems. As we explained earlier, different models have different positions regarding this capability. For example, the heuristic and biases program (Kahneman & Frederick, 2002) and some other DI models (Epstein, 1994; Evans, 2016; Sloman, 1996, but see Evans, 2017, 2018; Sloman, 2014, Stanovich, 2018) do not incorporate the existence of logical intuition into their conceptualizations of human reasoning. Contrary to these theories are those which consider intuitive logic as a central component of their models (De Neys, 2012; Handley & Trippas, 2015; Mata et al., 2017; Pennycook et al., 2015; Thompson & Johnson, 2014). Thus, this study is a further evaluation of these inconsistent models regarding the existence of logical intuition.

Considering the novelty of the concept of logical intuition in the reasoning and decision-making field, the underlying mechanisms of this capability is still unclear. To further our understanding of this concept, the second aim of the current study is to examine individual differences regarding this sensitivity to logic. Especially, we are trying to answer whether this sensitivity is independent of CA and ACS, or higher ability people have a better and more efficient logical intuition? The theoretical divergence also exists regarding the relationship between processing types and individual differences. On the one hand, some models propose that individual differences do not arise until the last stage of the reasoning process (Betsch & Glöckner, 2010; Darlow & Sloman, 2010; De Neys & Bonnefon, 2013; Morsanyi & Handley, 2012a; Stanovich & West, 2008b; Toplak et al., 2014). On the other hand, according to different theorists, high and low ability reasoners diverge in their T1 thinking (Frey, Johnson, & De Neys, 2018; Pennycook et al., 2012; Thompson & Johnson, 2014; Thompson et al., 2018) or even at the pre-reasoning stage (Mata et al., 2017).

To evaluate these alternative views, we conducted two separate experiments. In the first experiment, we used three different reasoning problems (i.e., syllogisms, conditionals, and disjunctions) as well as a CA measure. The first experiment comprised of a liking judgment task in which participants were instructed to rate their liking of the conclusion to each reasoning problem and a validity judgment task in which they were asked to judge the validity of those problems. In the second experiment, we asked participants to judge the brightness of the conclusion to logical arguments (brightness judgments task) and then, in the second block, rate their liking of the conclusions of those arguments (liking judgments task). Liking judgment and brightness judgment tasks have been used successfully in previous research as indices of logical intuition (Morsanyi & Handley, 2012a; Nakamura & Kawaguchi, 2016; Trippas et al., 2016). In the second experiment, in order to investigate individual differences in both ability and willingness to think analytically, participants also completed a measure of ACS as well as a CA test.

Chapter 2: Experiment 1- Liking and Validity Judgments

Introduction

Previous research has found that people liked valid and believable conclusions to reasoning problems more than invalid and unbelievable ones and this effect is independent of deliberative T2 thinking (Morsanyi & Handley, 2012b; Trippas et al., 2016). According to the fluency-affect intuition model (Topolinski, 2018), the semantic and logical coherency of valid and believable problems lead to a feeling of fluency, which in turn, results in a positive affect. Reasoners use this positive feeling and their feeling of rightness as cues to intuitive judgment (Thompson et al., 2013; Thompson & Morsanyi, 2012). In this experiment, we used a similar liking judgment measure to examine the intuitive sensitivity to logical structure. Participants were also instructed to judge the validity of those problems in the second part of the experiment. Based on the theoretical frameworks and empirical findings, we hypothesized that reasoners would rate valid and believable conclusions more likable than their invalid and unbelievable counterparts.

To investigate individual differences in intuitive logic, participants completed a verbal and quantitative reasoning task designed to measure CA. According to the late divergence accounts (Betsch & Glöckner, 2010; Darlow & Sloman, 2010; De Neys & Bonnefon, 2013; Morsanyi & Handley, 2012a; Stanovich & West, 2008b; Toplak et al., 2014), there should be no relationship between liking ratings of valid problems and CA. However, the early divergence accounts (Thompson & Johnson, 2014; Thompson et al., 2018) would predict that people with higher ability have a greater logic effect compared to those with lower CA.

Method

Participants

47 students from Macquarie University participated in this experiment (39 were female and 8 were male, Mean age= 22.45, SD= 7.03). Participants were fluent English speakers without any education in formal logic. Subjects received either two course credits or AU$ 15 for their participation in the experiment. Each session lasted approximately 45 minutes.

Design and Materials

The design of the experiment was a 2 (logic: valid vs. invalid) × 2 (belief: believable vs. unbelievable) × 3 (problem type: syllogisms vs. conditionals vs. disjunctions) × 2 (CA: high vs. low groups) with the first three variables manipulated in a repeated measure design.

Reasoning materials: We used 72 reasoning problems which included 24 conditionals (12 modus ponens and 12 modus tollens), 24 disjunctions (12 affirmations and 12 denials), and 24 single-model syllogisms. Each subtype had the same number of problems in validity by believability cells (i.e., valid believable [VB], valid unbelievable [VU], invalid believable [IB], invalid unbelievable [IU]). Valid modus ponens problems were in the form of "if p then q, p, therefore q" and their invalid counterparts were in the form of "if p then q, p, therefore not q." The valid form for modus tollens was "if p then q, not q, therefore not p" and the invalid form of these problems was "if p then q, not q, therefore p." The structure of the valid affirmation disjunction was "p or q, p, therefore not q," while the structure of the invalid problem of this type was "p or q, p, therefore q." The valid form of the denial disjunction has "p or q, not p, therefore q" structure and its invalid form has "p or q, not p, therefore not q" structure. The p and q positions in both subtypes of the disjunctions were randomized for each problem to minimize the order effect of premise propositions ("either monkeys are primates, or they are rodents" vs. "either monkeys are rodents, or they are primates"). Finally, we had the same number of syllogisms in each validity by believability condition. These syllogisms were selected in a way, such that we had 12 conclusions with A-C conclusion's direction (Some A are C) and 12 conclusions with C-A conclusions' direction (No C are A). Moreover, half of the conclusions had "some" quantifier (Some C are A) and the other half had "no" quantifiers (No A are C), and this condition is held for each of the validity by believability conditions (i.e., VB, VU, IB, IU). Except for changing a few contents to culturally familiar ones for the Australian participants, all of the problem contents and structures were adapted from Trippas et al. (Trippas et al., 2016). An illustration of the materials can be found in Table 1 (for a full list of the materials, see Appendices 1-3).

Some studies have demonstrated that superficial features such as the content of problems can affect participants' judgment in designs with the pre-determined fixed contents (Klauer & Singmann, 2013; Singmann et al., 2014). To avoid any confounding effect of content, we randomized the content across each logic by belief cell for each participant anew. For the conditionals, we created a set of 4 different problems with the same content but with different validity and believability status (i.e., VB, VU, IB, IU) for each argument. Each problem was selected randomly from its related set. This randomization was constrained in a way that each participant was presented with the same number of problems in each logic by belief cell for the total of 12 modus ponens and 12 modus tollens conditionals. This way, each participant was presented with a unique list of problems without any association between their validity status and contents. The same randomization procedure was performed for affirmations and denials disjunctions, with the additional randomization of the "p" and "q" order in the first premise.

Table 1 Different problem types across validity and believability conditions that were used in experiments 1 and 2

	Valid		Invalid	
Problem Type	Believable	Unbelievable	Believable	Unbelievable
Conditionals				
Modus ponens	P1: If it is winter [p] then it is freezing outside [q] P2: It is winter [p] C: It is freezing outside [q]	P1: If it is winter [p] then it is hot outside [q] P2: It is winter [p] C: It is hot outside [q]	P1: If it is winter [p] then it is hot outside [q] P2: It is winter [p] C: It is freezing outside [¬q]	P1: If it is winter [p] then it is freezing outside [q] P2: It is winter [p] C: It is hot outside [¬q]
Modus tollens	P1: If it is summer [p] then it is hot outside [q] P2: It is freezing outside [¬q] C: It is winter [¬p]	P1: If it is winter [p] then it is hot outside [q] P2: It is freezing outside [¬q] C: It is summer [¬p]	P1: If it is winter [p] then it is hot outside [q] P2: It is freezing outside [¬q] C: It is winter [p]	P1: If it is summer [p] then it is hot outside [q] P2: It is freezing outside [¬q] C: It is summer [p]
Disjunctions				
Affirmation	P1: Either the sun is yellow [p] or it is blue [q] P2: The sun is yellow [p] C: The sun is not blue [¬p]	P1: Either the sun is yellow [p] or it is blue [q] P2: The sun is blue [q] C: The sun is not yellow [¬p]	P1: Either the sun is yellow [p] or it is blue [q] P2: The sun is blue [q] C: The sun is yellow [p]	P1: Either the sun is yellow [p] or it is blue [q] P2: The sun is yellow [p] C: The sun is blue [q]
Denial	P1: Either the sun is yellow [p] or it is blue [q] P2: The sun is not blue [¬q] C: The sun is yellow [p]	P1: Either the sun is yellow [p] or it is blue [q] The sun is not yellow [¬p] The sun is blue [q]	P1: Either the sun is yellow [p] or it is blue [q] The sun is not yellow [¬p] The sun is not blue [¬q]	P1: Either the sun is yellow [p] or it is blue [q] The sun is not blue [¬q] The sun is not yellow [¬p]
Syllogisms				
Single-Model Syllogisms	P1: No burtes are marsupials P2: All bees are burtes C: No marsupials are bees	P1: No burtes are marsupials P2: All kangaroos are burtes C: No marsupials are kangaroos	P1: No burtes are marsupials P2: All kangaroos are burtes C: Some marsupials are kangaroos	P1: No burtes are marsupials P2: All bees are burtes C: Some marsupials are bees

To randomly assign the contents to syllogistic structures, we combined 24 categories (e.g., tools) with their members (e.g., hammer) or non-members (e.g., banana). For conclusions with the "no" quantifiers, believable conclusions were constructed by randomly pairing a category with its non-member (e.g., "no marsupials are parrots") and unbelievable conclusions were created by the combination of a category with one of its members (e.g., "no marsupials are kangaroos"). This pattern was reversed for conclusions featured with the "some" quantifiers. Combining a category with its member creates a believable conclusion ("some marsupials are kangaroos") and combining a category with its non-member results in an unbelievable conclusion ("some marsupials are parrots"). The middle "b" terms in the syllogistic structure are selected from a set of pseudowords (e.g., "firtes") to control the premises believability. The full set of category-member pairs and pseudowords can be found in Appendix 4.

Individual differences measure: Reasoners' CA was measured with the first part of the AH4 general intelligence test (Heim, 1970). This part contains 65 verbal and mathematical reasoning questions. Participants were given 10 minutes to answer as many questions as they could. Previous research has shown that this test can predict the performance across different reasoning tasks including syllogisms and conditionals (Newstead et al., 2004). An example of each verbal and mathematical questions is as listed below:

- *Here are three figures: 312. Divide the biggest figure by the smallest and add the result to the figure printed immediately after the smallest figure.*
- *When is to where as time is to . . . how / why / space / length / relativity*

Procedure

After signing a consent form, each participant completed the test on individual computers in small groups. The reasoning tasks were designed using PsychoPy 1.90.3 (Peirce et al., 2019; Peirce, 2008) and the CA task was created by Qualtrics (2013).

The experiment had three separate blocks. In the first block, participants were asked to rate their liking of the conclusion of reasoning problems. We used a serial presentation format, according to which, the first premise was presented for 2.5 seconds and removed from the screen. Then, the second premise was shown for 2.5 seconds and vanished after this period. Finally, the conclusion was presented along with a question ("how much do you like this statement") and a rating scale from 1 to 6. After rating their liking of all problems, participants performed a short memory recognition test that they were informed of at the beginning of the instructions. We asked participants to decide whether they had seen each of 20 words in the previous part. By using this test, we encouraged participants to read all the statements and not just the conclusion. In order to encourage participants to base their liking judgments on their gut feeling and thus, have a reliable index of logical intuition, there was no reference to logical reasoning in the instructions. The instructions were adapted from Trippas et al. (2016).

The instructions for the liking judgment part were as follows:

"In this part, we are interested in how much you like various statements. You will repeatedly be presented with three sentences in succession for a short amount of time. Please read these sentences and indicate how much you like the final sentence on a scale from 1 (dislike it very much) to 6 (like it very much) by clicking on a scale similar to the scale below [a screenshot of the rating scale was shown here]. When you make your liking judgment, rely on your intuition and feelings. Do not think about why you like or dislike the statement. Just go with your intuition and gut-feelings and do it as quickly as you can.

At the end of this part, you will be presented with a short memory task and will be asked to select those words that you have been presented during this part, so it is important that you pay attention to the sentences presented even though you are asked to respond based on your feelings. Please try to use the full extent of the scale (so use all possible values from 1 – 6). Before starting this section, you will be presented with three practice items."

In the second block of the experiment, participants were given the same reasoning problems and were asked to judge the logical validity of each problem. Since completing the validity task first may have encouraged participants to base their subsequent liking judgments on an explicit evaluation of problem validity, it was impossible to counterbalance the order of each block. However, to minimize the impact of liking judgments on subsequent validity judgments, we randomized content across logical structure again for this block. This way, we removed any systematic link between content and structure in both blocks of the experiment.

The presentation format for the second block of problems was similar to the first part of the experiment. The instructions for the validity judgment part were as stated below:

"In the second part of the experiment you will be presented with some statements again. Your task is to decide whether the conclusion (the last statement) necessarily follows from the premises (the first two statements). In this part, you have to reason on the basis of the logical validity. So, even if the statements don't mean anything or they are not true in real life, you

should assume they are true. A conclusion that necessarily follows is one that must be true (valid), assuming the premises are true.

You need to take the point of view of a perfectly logical person and only consider the information that is given to you, whether it makes sense or not. Think about your answer carefully. Do not rush. Choose your answer by clicking on the YES or NO boxes by moving the mouse. If you think the conclusion necessarily follows from the premises (i.e., it is valid) click on YES, otherwise click on NO. Before starting this section, you will be presented with three practice items. Click "NEXT" to start the practice trial."

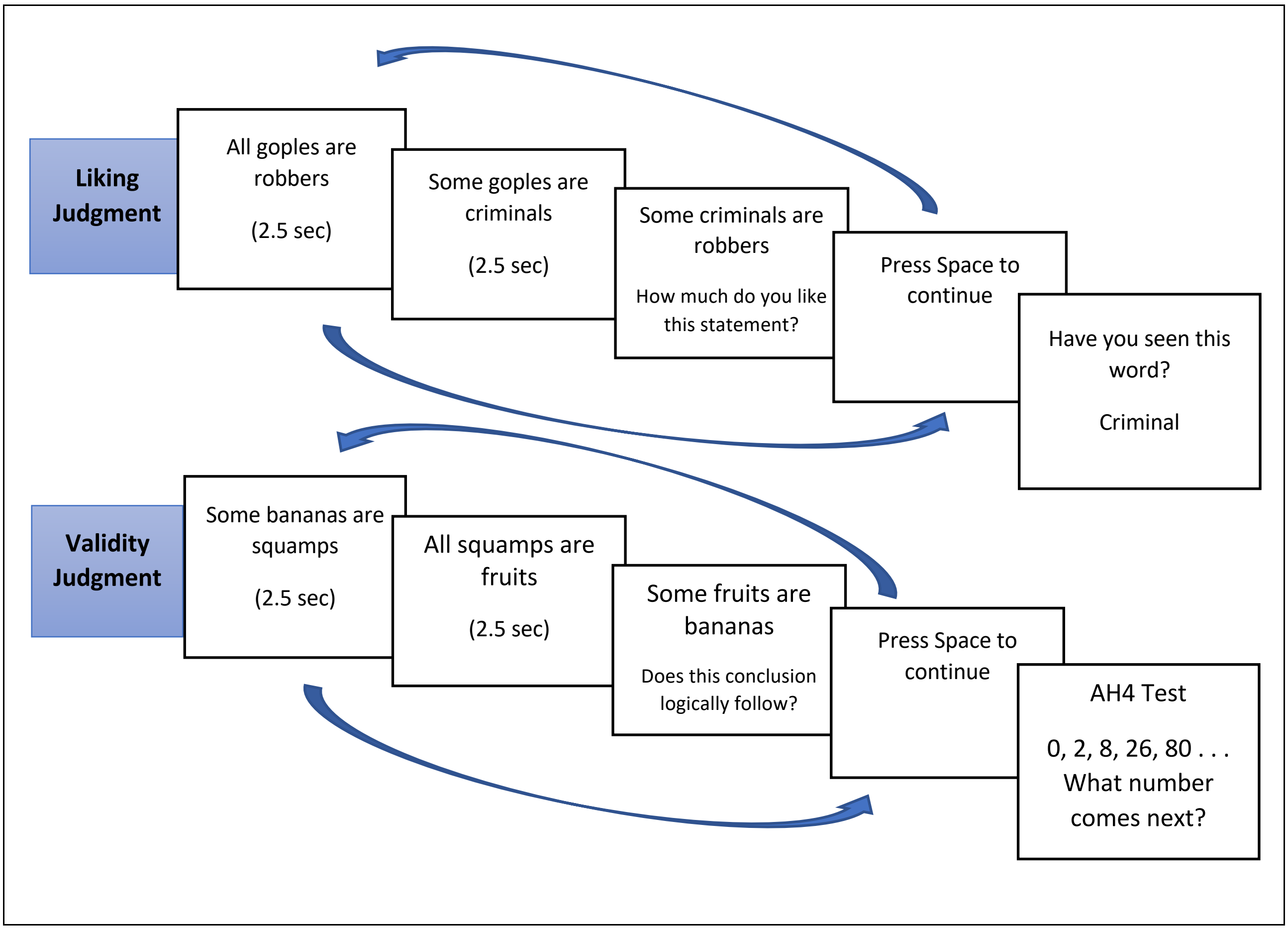

Figure 2 An overview of the experiment procedure

Before the actual test, participants solved 3 example problems for each part of the experiment. Upon the completion of the liking judgment and validity judgment parts, participants were presented with the CA test. They were instructed to solve 12 practice questions in an unlimited time and the actual test within 10 minutes. At the end of the experiment, participants were debriefed and thanked for their participation in the study. An illustration of the experimental procedure is presented in Figure 2.

Results

In this experiment, we aimed to answer two related questions. First, do participants show an intuitive logic effect as reflected in liking the conclusion on valid arguments more than invalid ones? If so, do variations in this capability relate to CA? In order to answer each question, we performed different statistical tests. We used a mixed analysis of variance (ANOVA) to examine the existence of logical intuition. For answering the individual differences question, we investigated the interactions in the mixed analysis of variance (ANOVA) as well as using several linear and multiple regression analyses.

Experimental Analysis

Following from Trippas, Handley, and Verde (2013), we performed a median split on AH4 scores. Participants were divided into two groups of high and low CA based on the median score of 43 on this test. Following this procedure, the low CA group included 25 participants (M= 36.84, SD= 4.95) and the high CA group comprised of 22 participants (M= 50.0, SD= 5.28). We analyzed two parts of the experiment separately. For the liking judgment part, participants with a memory score lower than two standard deviations from the mean of the total sample were excluded which include three participants. However, the key findings were not significantly different when all participants were analyzed. The analysis of the validity judgment task was performed on the total sample, including participants with the lowest memory score.

Liking Judgment

The first part of the analysis was performed on the liking rating scores. A 2 × 2 × 3 × 2 mixed analysis of variance (ANOVA) with logic (valid vs. invalid), belief (believable vs. unbelievable), and problem type (syllogism vs. conditional vs. disjunction) as within-subject factors and CA (high CA vs. low CA) as a between-subject factor revealed a significant main effect of logic, $F(1,42) = 50.97$, $p < .001$, $\eta_p^2 = .55$, a significant main effect of belief, $F(1,42) = 74.58$, $p < .001$, $\eta_p^2 = .64$, and a significant main effect of problem type, $F(2, 84) = 12.92$, $p < .001$, $\eta_p^2 = .24$. These findings indicate that participants liked the conclusion of valid (M= 4.13, SD= .11) and believable (M= 3.95, SD= .09) problems more than the conclusions of invalid (M= 2.73, SD= .11) and unbelievable (M= 2.91, SD= .08) ones. Moreover, a Bonferroni post-hoc test showed that participants liked both conditionals (M= 3.59, SD= .07) and disjunctions (M= 3.50, SD= .07) more than syllogisms (M= 3.18, SD= .07; $p < .001$, $p = .003$, respectively). There was no difference between liking ratings of disjunctions and conditionals ($p = .67$). The liking rating means for each condition across three problem types can be found in Table 2.

The findings showed a significant logic by belief interaction, $F (1,42) = 6.99$, $p = .01$, $\eta_p^2 = .14$. Follow up analyses showed that the effect of logic was more pronounced for believable problems. However, for both valid and invalid problems, participants liked believable problems more than unbelievable ones (ps < .001, ds > .98). We also found significant interaction of logic by problem type, $F (2,84) = 5.55$, $p = .005$, $\eta_p^2 = .12$, and belief by problem type, $F (1.62, 68.98) = 6.70$, $p = .004$, $\eta_p^2 = .14$. The results revealed that the effect of logic was more pronounced on conditionals and disjunctions and the effect of belief was more pronounced on disjunctions.

Table 2 Liking rating means (SD) for each logic by belief condition across different problem types and cognitive ability

		Valid		Invalid	
		Believable	Unbelievable	Believable	Unbelievable
Low CA	Syllogism	4.30 (.20)	2.94 (.25)	3.51 (.22)	2.36 (.16)
	Conditional	4.63 (.19)	3.66 (.19)	3.55 (.22)	2.60 (.19)
	Disjunction	5.20 (.17)	3.17 (.29)	3.67 (.27)	2.22 (.18)
	Total	**4.71 (.14)**	**3.26 (.20)**	**3.57 (.21)**	**2.40 (.14)**
High CA	Syllogism	4.15 (.22)	3.58 (.28)	2.61 (.24)	2.14 (.17)
	Conditional	4.94 (.21)	4.08 (.21)	2.94 (.24)	2.35 (.21)
	Disjunction	5.13 (.18)	3.79 (.32)	2.83 (.30)	2.02 (.20)
	Total	**4.74 (.15)**	**3.81 (.21)**	**2.79 (.23)**	**2.17 (.15)**
Total	Syllogism	4.22 (.15)	3.26 (.19)	3.06 (.16)	2.25 (.12)
	Conditional	4.78 (.14)	3.87 (.14)	3.25 (.16)	2.48 (.14)
	Disjunction	5.16 (.13)	3.48 (.21)	3.25 (.20)	2.12 (.14)
	Total	**4.72 (.10)**	**3.53 (.15)**	**3.18 (.16)**	**2.28 (.10)**

More importantly, the findings revealed a significant logic by CA interaction, $F (1,42) = 4.17$, $p = .047$, $\eta_p^2 = .09$. As Table 2 shows, differences between high and low CA groups were more pronounced on invalid problems, $t (42) = 2.21$, $p = .033$, $d = .66$. Compared to the low CA participants (M= 2.99, SD= .73), participants with higher CA (M= 2.48, SD= .78) rated invalid problems less likable. Higher CA participants (M= 4.28, SD= .79) rated valid problems more likable compared to low CA participants (M= 3.98, SD= .65), however, this difference did not reach to significance ($p = .18$). Moreover, both low and high CA groups liked valid problems significantly more than invalid ones (for high CA group, $t (19) = 5.70$, $p < .001$, d=

1.38, and for low CA group, t (23) = 4.12, $p < .001$, d= .84). Nevertheless, the difference between liking ratings for valid and invalid problems was larger for high CA participants compared to low CA participants (1.8 vs. 1.0).

There was also a significant belief by CA interaction, F (1,42) = 5.05, p = .03, η_p^2= .11. Participants with low CA (M= 4.14, SD= .64), compared to the participants with high CA (M= 3.77, SD= .44), rated believable problem more likable, t (42) = 2.21, p= .033, d= .67. Applying a Bonferroni correction to adjust the p value render this value non-significant. While high CA participants (M= 2.99, SD= .47) rated unbelievable problems more likable compared to the low CA participants (M= 2.82, SD= .53), this difference was not significant (p= .276). Moreover, both low and high ability groups liked believable problems significantly more than unbelievable ones (for high CA group, t (19) = 5.80, $p < .001$, d= 1.71, and for low CA group, t (23) = 6.90, $p < .001$, d= 1.75). However, the difference between liking ratings for believable and unbelievable problems was larger for low CA group compared to high CA group (1.31 vs. 0.78).

The key finding of this part of the analysis was the different liking rating for valid and invalid problem types. This finding revealed that reasoners consider the underlying logical structure of a problem even when they are not instructed to do so. Moreover, this effect was greater for participants with higher CA, compared to those with low CA. These results were consistent with the predictions of recent dual process theories and previous studies which investigated logical intuition by using different experimental design (Bago & De Neys, 2018; Bago et al., 2019a; Morsanyi & Handley, 2012a; Thompson & Johnson, 2014; Trippas et al., 2016). We discuss the implication of these findings in more details in the discussion section.

Validity judgments

A 2 (Logic: valid vs. invalid) × 2 (Belief: believable vs. unbelievable) × 3 (Problem type: syllogism vs. conditional vs. disjunction) × 2 (CA: high CA vs. low CA) mixed analysis of variance (ANOVA) revealed a significant main effect of logic, F (1,45) = 336.60, $p < .001$, η_p^2 = .88, indicating that participants endorsed valid problems (M= .84, SD= .02) more than invalid problems (M= .18, SD= .03). The main effects of belief and problem type were not significant (p = .21, p = .08, respectively) which is as expected considering the simple nature of the problems. Participants endorsed believable problems (M= .52, SD= .02) approximately at the same rate to the unbelievable problems (M= .50, SD= .02). The results showed a significant difference between total endorsement rating of Low CA (M= .55, SD= .02) and High CA groups (M= .48, SD= .02), F (1,45) = 7.98, p = .007, η_p^2= .15.

The results revealed a logic by belief interaction, F (1,45) = 12.18, p = .001, η_p^2 = .21, which indicates that the effect of logic endorsement was more pronounced for believable problems. The average endorsement rate for each condition across three problem types can be found in Table 3.

Table 3 Average endorsement rates (SDs) of validity judgment part for each cognitive ability group

		Valid		Invalid	
		Believable	Unbelievable	Believable	Unbelievable
Low CA	Syllogism	0.84 (0.04)	0.77 (0.05)	0.28 (0.05)	0.28 (0.04)
	Conditional	0.89 (0.03)	0.77 (0.04)	0.23 (0.04)	0.25 (0.04)
	Disjunction	0.89 (0.04)	0.84 (0.05)	0.25 (0.05)	0.24 (0.05)
	Total	**0.87 (0.02)**	**0.80 (0.03)**	**0.26 (0.04)**	**0.26 (0.03)**
High CA	Syllogism	0.88 (0.04)	0.84 (0.06)	0.15 (0.05)	0.13 (0.04)
	Conditional	0.82 (0.04)	0.77 (0.04)	0.05 (0.05)	0.12 (0.05)
	Disjunction	0.90 (0.04)	0.89 (0.05)	0.09 (0.06)	0.07 (0.05)
	Total	**0.87 (0.03)**	**0.83 (0.04)**	**0.10 (0.04)**	**0.11 (0.04)**
Total	Syllogism	0.86 (0.03)	0.81 (0.04)	0.22 (0.03)	0.21 (0.03)
	Conditional	0.85 (0.03)	0.77 (0.03)	0.14 (0.03)	0.19 (0.03)
	Disjunction	0.89 (0.03)	0.86 (0.04)	0.17 (0.04)	0.16 (0.04)
	Total	**0.86 (0.02)**	**0.81 (0.03)**	**0.18 (0.03)**	**0.18 (0.03)**

Similar to the liking judgment task, the results of the validity judgment task yielded a logic by CA interaction, F (1,45) = 5.69, p = .021, η_p^2 = .11. As is apparent from Table 3, the difference between high (M= .10, SD= .09) and low CA groups (M= .26, SD= .21) was more pronounced on invalid problems, t (45) = 3.34, p= .002, d= .99. These groups did not show any significant differences regarding valid problems (p= .68). In other words, higher ability participants endorsed invalid problems less, compared to the low ability participants. Moreover, both low and high ability groups endorsed valid problems significantly more than invalid ones (for high CA group, t (21) = 17.49, p < .001, d= 3.81, and for low CA group, t (24) = 10.23, p < .001, d= 2.04). However, the difference between valid and invalid endorsement ratings was higher for the high ability group compared to the low ability group (0.75 vs. 0.57). There was no significant interaction between belief and CA (p > .55). A summary of the results for both parts of the experiment can be found in Appendices 7.

Individual Differences

The analyses in the previous section concerning CA were based on a common practice of the median split of a continuous variable (Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015). Although using the CA as a covariate in an analysis of covariance (ANCOVA) revealed no significant changes in the results of the ANOVA, in order to avoid the possible problems caused by artificial categorization of a variable (McClelland, Lynch, Irwin, Spiller, & Fitzsimons, 2015; Rucker, McShane, & Preacher, 2015), we replicated the ANOVA findings by using linear regression.

Before analyzing the results for each instructional condition, checking the correlation between the variables of the experiment can illuminate the general trends of the data. To investigate these relationships, different logic, belief, and logic by belief interaction indices were created for each part of the experiment. These indices were calculated as follows:

Logic Index = VB + VU – IB – IU

Belief Index = VB + IB – VU – IU

Interaction Index = VU + IB – VB – IU

Table 4 shows the correlation coefficients between these indices and CA for both parts of the experiment. Logic indices in the liking judgment task and the validity judgment task were positively correlated. There were also positive relationships between each of these indices and CA. On the other hand, while belief indices in the first and second parts of the experiments were positively correlated, they had negative relationships with both logic indices and CA.

Table 4 Correlation coefficients between CA and logic and belief indices in liking judgment and endorsement judgment parts of the first experiment

		1	2	3	4	5	6	7
1	Logic Liking Index	1						
2	Belief Liking Index	-.53 **	1					
3	Interaction Liking Index	-.69 **	.002	1				
4	Logic Endorsing Index	.49 **	-.27	-.34 *	1			
5	Belief Endorsing Index	-.26	.41 **	.06	-.41 **	1		
6	Interaction Endorsing Index	-.002	-.05	-.02	.11	.001	1	
7	Cognitive Ability	.47 **	-.30 *	-.31 *	.47 **	-.28	.12	1

Note: ** means correlation is significant at the 0.01 level and * is significant at the 0.05 level

Several simple linear regressions with CA and logic index for validity judgment as predictors and logic effect for liking judgment as outcome showed that CA and validity logic index are able to predict the logic effect in liking judgment, β (CA) = .451, R^2 = .203, F (1, 42) = 10.70, p = .002, and β (validity logic index) = .489, R^2 = .239, F (1, 42) = 13.17, p = .001. However, as Table 5 shows, after entering both variable simultaneously into a regression model, only the validity logic index remains a significant predictor of liking effect. The multiple regression analysis revealed that after controlling for the explicit measure of reasoning, CA could not predict liking judgment ratings.

Table 5 Simple linear regression analyses predicting logic and belief indexes with CA in both parts of the experiment

	Predictor	β	t	p	F	R^2	P (F)
Liking Judgment Task	CA	.283	1.91	.063			
	Validity logic index	.355	2.39	.021			
					8.82	.301	.001
Validity Judgment Task	CA	.308	2.18	.035			
	Liking logic index	.349	2.47	.017			
					10.20	.317	.000

On the other hand, both CA and liking logic index could independently predict the logic effect in the endorsement judgment task β (CA) = .471, R^2 = .222, F (1, 45) = 12.85, p = .001, and β (liking logic index) = .493, R^2 = .243, F (1, 45) = 14.45, p < .001. More importantly, after controlling the liking logic index, CA still significantly predicted the logic effect in the endorsement judgment task. We will discuss the implication of these results in the next section.

Discussion

Recent dual process theories argue that people can access the underlying logical structure of a problem relatively intuitively and automatically. To test this hypothesis, participants were asked to rate their liking of the conclusion to several reasoning problems. To make sure that no superficial features render the effects, each argument was randomly assigned to a validity by believability cell. Moreover, to encourage participants to base their judgments on their T1 thinking, there was no sign cueing the logical judgment in the instructions and participants were asked to take their gut feeling and affects into account in rating their liking. Nevertheless, the results showed that subjects rated the conclusions to valid and believable arguments more likable than those to invalid and unbelievable arguments.

These results are consistent with the findings of experiments conducted by Morsanyi and Handley (2012), Trippas and colleagues (2016), and Nakamura and Kawaguchi (2016). According to these researchers, the underlying logical structure of the valid problems and its coherency create a sense of fluency. Despite being unaware of this conceptual fluency, reasoners find it desirable, which in result, leads to a positive affect. This final affective product, then, will be used as a criterion in liking judgments (Topolinski, 2011; Topolinski & Strack, 2009). The same account would hold for believable problems.

Moreover, the results revealed that reasoners show individual differences regarding the logic effect in the liking judgment task. The findings indicate that smarter participants show a greater logical intuition capacity as is reflected in their liking judgments. While there were no significant differences in liking ratings of valid problems for both ability groups, higher CA participants, compared to their low CA counterparts, rated invalid problems less likable. However, after controlling for explicit reasoning measures, CA could no longer predict the differences in intuitive measure. It seems probable that, in rating their liking judgment, participants may rely on their deliberative thinking. In the absence of any obvious and definitive features to base their liking judgment on, reasoners might evaluate the believability and validity of those statements as their criteria for the liking judgment. In other words, by engaging in deliberative thinking, it is possible that participants based their judgment on the validity of the statements instead of their own gut feelings or affective states. Thus, considering these findings, one can argue that the deliberative components of liking judgment, and not the intuitive ones, renders such individual differences in judgment. These findings support the late emergence of individual differences in the reasoning process (Betsch & Glöckner, 2010; Darlow & Sloman, 2010; De Neys & Bonnefon, 2013; Morsanyi & Handley, 2012a; Stanovich & West, 2008b; Toplak et al., 2014). According to these theories, although the compatibility of a final response with normative answers can in part happen at the beginning of the reasoning process, reasoners with different levels of CA have a similar capability at reaching the normative answer intuitively. Our findings indicate that reasoners, at least those with higher or lower CA scores, possess similar intuitive capabilities.

Chapter 3: Experiment 2- Brightness and Liking Judgments

Introduction

The findings of Experiment 1 support the claim that people possess an intuitive sensitivity to logic. Participants liked the conclusion to valid arguments significantly more than invalid arguments. Besides the content and structure of the statement, there were no external cues to signal liking judgment. Since the content of the problems varied for each participant, it can be argued that the logical structure of the problems affected liking ratings, rather than a superficial feature linked to problem content. However, it is still unclear how the underlying structure influences liking ratings. One possible claim, as we discussed in the previous section, is that participants may engage in deliberative thinking while rating their liking. Previous findings have addressed this critique in different ways. Morsanyi and Handley (2012) compared the working memory capacity of participants as an index of T2 thinking efficiency on the liking judgment and the validity judgment tasks. They found that while participants with higher working memory had a better performance in the validity judgment part, there was no difference between high and low working memory group regarding their liking rating. This finding indicates that while liking judgment is independent of deliberative T2 thinking, participants reveal no individual differences regarding this intuitive measure.

Another method to measure logical intuition is to use a purer measure of logical intuition which provides an external criterion for judgment. For example, Trippas et al. (2016) manipulated the contrast of each problem and asked participants to judge the brightness of those statements. According to these researchers, a brightness judgment task is entirely unrelated to an explicit reasoning judgment. The rationale of this task is similar to the liking rating task. Coherency can lead to a brightness judgment the same way as it does in liking judgment; through fluency and then positive affect. By using this method, these researchers found that reasoners judged valid problems as brighter than their invalid counterparts. In this experiment, we used brightness judgment as well as the liking judgment task.

Another issue regarding the first experiment is the individual differences measure. According to Stanovich (2018), intelligence tests, including the AH4 that we used in the previous experiment, are among optimal performance tasks. In these tasks, participants were asked to show their best performance and try to solve questions as accurately as they can. On the other hand, this is not similar to the real-world situations where there is not a signal for optimal responding. Normal performance tasks lack any cue to optimal performance. One example of such tasks is Analytic Cognitive Style (ACS) tasks in which the structural features of the problem suggest a compelling and initial but wrong answer that needs to be overridden to reach a correct answer. As Stanovich (2018) proposed, conflict detection, as an index of logical intuition, is more related to thinking disposition and style rather than CA. Thus, in order to examine individual differences in more depth, we included an ACS measure in the second experiment in addition to CA.

We predicted that participants would rate the conclusion to valid and believable arguments to be brighter than those to invalid and unbelievable ones. Moreover, brightness judgments provide us with a further test of the individual differences in intuitive logic. Replicating the findings of Experiment 1 with the brightness judgment task regarding the lack of individual differences for participants with different cognitive ability and styles would support the late divergence accounts of reasoning process.

Method

Participants

47 undergraduate students (35 were female, and 12 were male, Mean age= 21.45, SD= 7.40) were recruited from SONA participation pool at Macquarie University. Participants were fluent English speakers who took part in the experiment in exchange for either two course credits or AU$ 15. Each session lasted for 55 minutes.

Design and Materials

The design of this experiment was a 2 (logic: valid vs. invalid) × 2 (belief: believable vs. unbelievable) × 3 (problem type: syllogisms vs. conditionals vs. disjunctions) × 2 (brightness: high vs. low brightness) × 2 (CA: high vs. low groups) × 2 (ACS: high vs. low groups) with the first four variables manipulated in a repeated measure format.

Reasoning materials: In order to avoid any confounding effect of physical brightness with the logical structure of the problems, we randomly assigned the contrast value for each item, independent of its logical structure or content. To achieve this goal, we increased the number of reasoning materials to 96 problems; including 32 conditionals (16 modus ponens and 16 modus tollens), 32 disjunctions (16 affirmations and 16 denials), and 32 simple syllogisms. This way, we ended up with the same number of high and low contrast items in each validity by believability cell without disarranging the counterbalanced format of the reasoning materials in the first experiment. As an example, consider 16 modus ponens conditionals. For this subtype, we had 4 VB items, 4 VU items, 4 IB items, and 4 IU items and for each of these conditions, half of the problems had high brightness, and the other half had low brightness. This was the same for other problem types, including syllogisms. This randomization was conducted for each item independently and for each participant anew.

We used a similar manipulation and format for the physical brightness to those designed by Trippas et al. (2016). By changing the contrast of the black text against a white background, we created items with high and low brightness. High brightness items were created by randomly drawn an RGB (red, green, blue) value from a normal distribution with the mean of 30 and the standard deviation of 1. Low brightness items were also constructed by randomly selecting a value from a normal distribution with the mean of 40 and the standard deviation of 1. All elements of the problem, including premises and conclusion, had the same RGB value.

Finally, to preserve the exclusivity of category and member pairs due to the addition of 8 new syllogisms, we created 16 categories each with 2 members. The pairing method of categories and members for conclusions with "some" and "no" quantifiers remained the same as the previous experiment. Besides these modifications, all the materials, structures, contents, and randomization methods were similar to Experiment 1. For a full set of categories and members, see Appendix 5.

Individual differences measures: We used part one of the AH4 intelligence task as a measure of CA. In this experiment, however, we added an ACS measure as the second index of individual differences. The fundamental feature of ACS items was cueing an intuitive but wrong answer upon the reading of the question (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014b; Trippas, Pennycook, Verde, & Handley, 2015). To respond normatively, one needs to detect the wrongness of this initial answer and override it. For example, consider the well-known bat-and-ball CRT item taken from Frederick (2005):

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

Participants, even those from the high-ranked universities, usually incorrectly respond 10 cents (Frederick, 2005). This answer is incorrect because if the ball is 10 cents and the bat is one dollar more than the ball, their sum is 1.20 cents which exceeds the 1.10 cents on the question. The willingness and the ability of participants can lead them to find out the correct answer is 5 cents.

To measure the ACS, we used 6 incongruent base-rate problems taken from De Neys and Glumicic (2008), 3-item CRT from Frederick (2005), and 4-item CRT designed by Toplak et al. (2014). In the base-rate problems, participants were presented with the proportion of two different groups in a sample of 1000 people followed by a short personality description of a randomly selected person from that sample. Participants were asked to which group this person more likely belongs. We used three extreme base rates (995/5; 996/4; 997/3) for every two problems. An instance of this problem is as follow. The total of 13 ACS questions can be found in Appendix 6.

In a study, 1000 people were tested. Among the participants, there were 5 engineers and 995 lawyers. Jack is a randomly chosen participant of this study.

Jack is 36 years old. He is not married and is somewhat introverted. He likes to spend his free time reading science fiction and writing computer programs.

What is most likely?
a. Jack is an engineer
b. Jack is a lawyer

Procedure

This experiment had a brightness judgment block, a liking judgment block, and an individual differences block. In the first block, participants were asked to judge the brightness of the conclusions to each reasoning problem (see Figure 3). The sequential presentation format and the timing of each premise was similar to the first experiment. After finishing the brightness task, participants performed a short memory recognition test in which they had to decide whether they had seen each of 20 words in the previous part. The instructions for the brightness judgment task were adapted from Trippas et al. (2016) and were as follows:

"You will repeatedly be presented with three sentences in succession for a short amount of time. Please read these sentences and indicate whether the final statement is high or low brightness

on a scale from 1 (certainly low brightness) to 6 (certainly high brightness) by clicking on a scale similar to the scale below [a screenshot of the rating scale was shown here]. The brightness differences are quite subtle, so make use of your gut feeling and intuition when making a judgment. Occasionally you have to guess. Once again, make sure you do this on the basis of your gut feeling and do it as quickly as you can. Also, make sure to use the full extent of the scale (so use all possible values from 1 – 6).

At the end of this part, you will be presented with a short memory test to check whether you read all the sentences, so it is important that you pay attention to the sentences presented even though you are asked to respond based on the brightness of each statement. Before starting this section, you will be presented with three practice items. Click "NEXT" to start the practice trial."

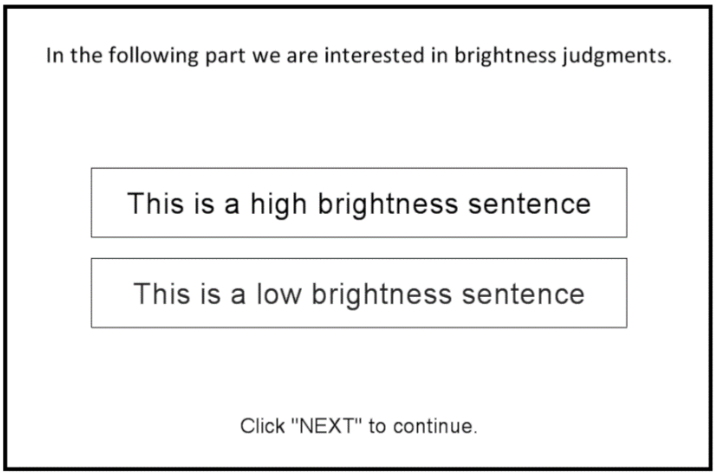

Figure 3 A screenshot of the instructions of the second experiment regarding the high and low brightness statements

At the second block, participants were given the same reasoning problems and were asked to rate their liking of each problem. At the end of this block, participants were presented with a similar memory test. Since we assumed that the brightness judgment provides a more reliable measure of logical intuition compared to liking judgment, the order of those tasks was not counterbalanced. However, to minimize the order effect, we independently allocated content across logical structure for the second block. This way, we ended up with different list of problems for the second block and removed any systematic link between content and structure in both blocks of the experiment. The presentation format for liking judgment task was similar to the first experiment.

Upon the completion of the brightness judgment and liking judgment task, participants were presented with the ACS and CA measures. They solved 13 ACS questions first, in an unlimited time and then were instructed to solve 12 practice questions and 65 actual questions of the CA test. At the end of the experiment, participants were debriefed and thanked for their participation in the study.

Results

As in Experiment 1, we present the experimental analysis followed by an individual differences analysis. In the first part, we used a mixed analysis of variance (ANOVA) to examine the existence of logical intuition. In the second part, we used several simple and multiple linear regression analyses to check the individual differences regarding intuitive logic.

Experimental Analysis

Following the first experiment, we excluded participants with a memory score lower than 2 standard deviations from the average score for each part of the experiment. This exclusion resulted in removing 1 participant from the first part and 2 participants from the second part of the analysis. Nevertheless, analyzing the whole sample yielded no significant changes in the key findings.

We also performed a median split on CA and ACS measures. The median score for CA measure was 43 which divided the sample into 22 high CA participants (M= 48.59, SD= 3.05) and 25 low CA participants (M= 35.72, SD= 5.44). Moreover, based on the median score of 3 on the ACS measure, two groups of 23 high ACS (M= 6.78, SD= 1.45) and 24 low ACS (M= 1.33, SD= 1.21) participants were created. High and low ACS are those with reflective and intuitive cognitive style, respectively (Trippas, Pennycook, Verde, & Handley, 2015).

Brightness Judgment

Since we had two between subject variable for individual differences, we performed two separate mixed analysis of variance (ANOVA) for each of CA and ACS measures. A 2 × 2 × 3 × 2 × 2 mixed analysis of variance (ANOVA) with logic (valid vs. invalid), belief (believable vs. unbelievable), problem type (syllogism vs. conditional vs. disjunction), and brightness (high vs. low contrast) as within-subject factors and CA (high CA vs. low CA) as between-subject factor yielded a significant main effect of logic , $F (1,44) = 10.37$, $p = .002$, $\eta_p^2 = .19$, a significant main effect of belief, $F(1,44) = 16.05$, $p < .001$, $\eta_p^2 = .27$, a significant main effect of problem type, $F(1.52, 66.73) = 3.55$, $p = .046$, $\eta_p^2 = .075$, and a significant main effect of brightness, $F(1,44) = 25.50$, $p < .001$, $\eta_p^2 = .37$. Participants judged the conclusion to valid (M= 4.06, SD= .09), believable (M= 4.07, SD= .08) and bright problems (M= 4.15, SD= .11) to be brighter than those to invalid (M= 3.68, SD= .12), unbelievable (M= 3.66, SD= .12), and less bright (M= 3.58, SD= .09) problems. Although there was an evidence of a main effect of problem types, the post-hoc tests did not reveal any significant difference between these problem types (M= 3.79, SD= .11 for syllogisms, M= 3.95, SD= .08 for conditionals, and M= 3.86, SD= .09 for disjunctions).

The results revealed a significant logic by belief interaction, $F(1,44) = 4.71$, $p = .035$, $\eta_p^2 = .097$. The effect of logic was more pronounced in believable condition (M difference of valid and invalid problems were 0.46 for believable condition and 0.30 for unbelievable condition), $t(45) = 3.78$, $p < .001$, $d = .47$ (see Table 6). We also found significant brightness by CA, $F(1,44) = 5.20$, $p = .027$, $\eta_p^2 = .106$, logic by problem types, $F(2,88) = 3.23$, $p = .044$, $\eta_p^2 = .07$, and brightness by problem types, $F(2,88) = 3.46$, $p = .036$, $\eta_p^2 = .07$, interactions. Follow up analyses revealed that the logic effect was more pronounced on conditionals and disjunctions, and brightness ratings were more marked for syllogisms and for participants with higher CA scores. Moreover, we found a significant difference between total brightness rating of Low CA (M= 4.06, SD= .12) and High CA groups (M= 3.68, SD= .13), $F(1,44) = 4.61$, $p = .037$, $\eta_p^2 = .095$.

Table 6 Average brightness rating for each condition (SDs) across cognitive ability and style groups

		Valid		Invalid	
		Believable	Unbelievable	Believable	Unbelievable
Cognitive Ability	Low Group	4.47 (.13)	4.17 (.16)	3.92 (.16)	3.68 (.21)
	High Group	4.14 (.14)	3.45 (.16)	3.77 (.16)	3.36 (.22)
Cognitive Style	Low Group	4.34 (.14)	3.92 (.17)	4.05 (.15)	3.69 (.21)
	High Group	4.28 (.14)	3.72 (.18)	3.63 (.16)	3.34 (.21)

More importantly, the findings revealed that neither the logic by CA interaction, $F(1,44) = 1.50$, $p = .227$ $\eta_p^2 = .03$, nor the logic by ACS interaction, $F(1,44) = 1.13$, $p = .293$, $\eta_p^2 = .03$, were significant. The lack of significant interaction with individual differences measures was also true for the belief effect ($ps > .173$). Finally, using ACS as between-subject variable in the ANOVA instead of CA did not change the key main effects and interactions.

Liking judgment part

We performed a 2 (Logic: valid vs. invalid) × 2 (Belief: believable vs. unbelievable) × 3 (Problem type: syllogism vs. conditional vs. disjunction) × 2 (CA: high CA vs. low CA) mixed analysis of variance (ANOVA). The results indicated that participants liked conclusions to valid problems (M= 3.84, SD= .09) more than those to invalid problems (M= 3.18, SD= .09), $F(1,43) = 27.48$, $p < .001$, $\eta_p^2 = .39$. They liked the conclusions of believable arguments (M= 4.11, SD= .08) more than the conclusions of unbelievable ones (M= 2.91, SD= .10), $F(1,43) = 77.04$, $p < .001$, $\eta_p^2 = .64$. The analysis revealed a significant main effect of problem Type, $F(1.59, 68.35) = 23.35$, $p < .001$, $\eta_p^2 = .35$. Bonferroni post-hoc test showed that participants liked both conditionals and disjunctions more than syllogisms (both ps < .001). There was no difference between liking ratings of disjunctions and conditionals.

Similar to the first experiment, the results showed a significant logic by CA interaction, $F(1,43) = 4.82$, $p = .034$, $\eta_p^2 = .10$. The difference between liking ratings of high (M=4.06, SD= .44) and low CA groups (M= 3.61, SD= .70) was more pronounced on valid problem, indicating that high CA subjects rated valid problems more likable compared to low CA subjects, $t(43) = -2.67$, $p= .015$, $d= .76$. High CA participants judged the conclusions to valid problems significantly more likable than those to invalid problems, $t(20) = 5.35$, $p>.001$, $d= 2.69$. Although this difference was significant for low CA participants (p=.043), it was larger for the high CA group compared to the low CA group (.93 vs. .38).

Table 7 Average liking ratings for each condition (SDs) across cognitive ability and style groups

		Valid		Invalid	
		Believable	Unbelievable	Believable	Unbelievable
Cognitive Ability	Low Group	4.22 (.13)	3.00 (.19)	3.79 (.16)	2.67 (.15)
	High Group	4.74 (.13)	3.38 (.20)	3.67 (.17)	2.60 (.16)
Cognitive Style	Low Group	4.45 (.14)	3.06 (.19)	3.65 (.16)	2.49 (.14)
	High Group	4.48 (.15)	3.31 (.20)	3.84 (.17)	2.81 (.15)

We also found significant logic by problem type, $F(2,86) = 5.78$, $p = .004$, $\eta_p^2 = .12$, and belief by problem type, $F(1.55, 66.63) = 10.14$, $p < .001$, $\eta_p^2 = .19$, interactions. The effect of logic was more marked for conditionals, while the effect of belief was more pronounced for disjunctions. Finally, using ACS as the between-subject variable in the analysis revealed that, in contrast to the CA, ACS did not interact with logic in the liking ratings task (p= .665). Similarly, we did not find belief by ACS (p= .524) and belief by CA (p= .853) interactions.

Moreover, using CA or ACS as continuous variables in the analysis of covariance (ANCOVA) yielded similar results with the ANOVA based on the median split. The average liking ratings for each logic by belief cell can be found in Table 7. A summary of the main effects and interactions for both parts of the experiment can be found in Appendices 7.

Individual Differences

Before discussing the regression analyses, we examined the correlation matrix in Table 6. The indices for logic, belief, and interaction were created in a similar method to Experiment 1. As Table 8 shows, logic index in the brightness judgment task has a positive but non-significant relationship with ACS and a negative non-significant relationship with CA. The logic index in the liking judgment task, however, correlated significantly and positively with CA, but not with ACS. Logic and belief indices in both parts of the experiment were correlated. There was a significant relationship between ACS and CA, that is consistent with other findings in the literature which have shown that these two measures are highly correlated (e.g., Thompson et al., 2018).

Table 8 Correlation coefficients between CA, ACS, and logic and belief indices in brightness judgment and liking judgment tasks of the second experiment

		1	2	3	4	5	6	7	8
1	Logic Brightness Index	1							
2	Belief Brightness Index	.24	1						
3	Interaction Bright Index	.034	-.34 *	1					
4	Logic Liking Index	.40 **	-.07	.32 *	1				
5	Belief Liking Index	-.11	.50 **	-.37 **	-.26	1			
6	Interaction Liking Index	.11	.03	.15	-.05	-.12	1		
7	Cognitive Ability	-.05	.13	-.08	.35 *	-.12	-.11	1	
8	Cognitive Style	.11	.02	-.16	-.02	-.20	.06	.45 **	1

Note: ** means correlation is significant at the 0.01 level and * is significant at the 0.05 level

As shown in Table 8, the logic index in the brightness ratings is not significantly correlated with CA and ACS. These findings indicate that using a different measure of logical intuition removes the relationship between individual differences in logical intuition and CA found in Experiment 1 and the second part of Experiment 2. In support of this finding, we performed several linear and multiple regression analyses for logic indices of the brightness and liking judgment task with CA, ACS and logic indexes as predictors.

In the brightness judgment task, our results revealed that neither CA nor ACS is a significant predictor of the logic effect in the brightness judgment task, β (CA) = -.49, R^2 = .002, F (1, 44) = .107, p = .745, and, β (ACS) = .117, R^2 = .014, F (1, 44) = .611, p = .439. On the other hand, higher liking logic scores significantly predicted the brightness effect, β = .399, R^2 = .159, F (1, 44) = 8.34, p = .006. Moreover, as Table 9 shows, including CA, ACS and liking logic index simultaneously in a regression model revealed that this model can significantly predict the logic effect in the brightness task, R^2 = .26, F (3, 42) = 4.913, p = .005. In other words, the results showed that liking logic index is a significant positive predictor of logic effect in the brightness task, but this relationship is not explained by CA or ACS. Contrary to our expectation, CA negatively predicted the logic effect in the brightness task.

Table 9 Multiple linear regression analyses predicting logic and belief judgment with CA and ACS

	Predictor	β	t	p	F	R^2	P (F)
Brightness Judgment Task	CA	-.360	-2.22	.032			
	ACS	.281	1.85	.072			
	Liking logic index	.524	3.63	.001			
					4.91	.26	.005
Liking Judgment Task	CA	.506	3.61	.001			
	ACS	-.290	-2.052	.047			
	Brightness logic index	.457	3.61	.001			
					7.81	.364	.000

The findings of the regression analyses for the logic effect in liking judgments revealed that, whilst ACS is not a significant predictor of logic, β= -.005, R^2 = .000, F (1, 43) = .001, p = .974, higher level of CA and higher brightness logic scores positively predicted the logic index on this task, β (CA) = .356, R^2 = .127, F (1, 43) = 6.26, p = .016, and, β (brightness logic index)= .398, R^2 = .158, F (1, 43) = 8.10, p = .007. As is presented in Table 9, both CA and the brightness logic index remained significant predictors of the liking effect even after entering all predictors simultaneously in a regression model. On the other hand, ACS predicted the liking effect negatively after controlling other variables. These findings were consistent with the significant predictive power of CA in the first experiment. Taken together, these results support the findings on the first experiment and are consistent with the findings that individual differences in T1 thinking may not be predicted by cognitive abilities.

Discussion

In rating their liking, one can argue that participants may engage in deliberative thinking and consider the logical structure of the statements. However, rating the physical brightness of a statement is far more unrelated to explicit reasoning. By using a brightness judgments task and a liking judgments task, we replicated the key findings of Experiment 1 regarding the existence of logical intuition. In other words, participants judged valid and believable conclusions of reasoning problems to be brighter and more likable than invalid and unbelievable conclusions. This indicates that, even in the absence of any external cues or any instructions encouraging explicit reasoning, participants are able to distinguish between valid and invalid arguments. These findings are consistent with research which show that participants can reach the correct normative responses by using their T1 thinking and without the engagement of the corrective T2 thinking (Bago & De Neys, 2018; De Neys & Bago, 2017a; Newman et al., 2017).

Interestingly, the individual differences analysis of brightness judgment provides a different picture to the analysis of liking judgments. Whilst we replicated the finding of Experiment 1 regarding the relationship between CA and logic effects in liking judgments, such differences were not found in the brightness judgment task. In other words, neither CA nor ACS could predict the logic effect in the brightness judgment task. It seems that using an arguably purer logical intuition measure rendered the logic and ability relationship non-significant. Based on this finding, reasoners with different ACS and CA have similar intuitive thinking as reflected in their brightness ratings of valid and invalid statements. However, the negative predictive value of CA in the brightness task, as shown by regression analyses, was surprising. One possibility is that higher CA participants may rely on different responding strategies than those intended to measure by the experiment. In other words, this group may base their response on a pure judgment of brightness without reading the structure carefully. No matter what kind of strategy they used, it is apparent that the effect of logic on brightness judgments is not different for reasoners with high and low cognitive abilities. If anything, this effect seems to arise when participants are less engaged in explicit reasoning or have lower CA scores.

On the other hand, the liking judgment findings revealed that reasoners with different CA show varying effects of logic. Controlling for the brightness effect did not change the predictive power of CA for the liking effect. In the next section, we will discuss these results in more detail and review the implications of these findings for the recent dual process accounts.

Chapter 4: General Discussion

According to some theoretical conceptualizations regarding how reason and intuition interact, access to the logical structure of problems requires primarily T2 thinking. This type of thinking is assumed to be responsible for monitoring the quality of the responses (Kahneman & Frederick, 2002, but see, Evans, 2017) or correcting initial heuristic responses if required (Evans & Stanovich, 2013). These frameworks have described T1 processes as a type of thinking which mainly operates on the superficial features of a problem (Kahneman & Frederick, 2002; Sloman, 1996). The present study aimed to investigate the accessibility of simple logical structure arising from T1 thinking and to examine the extent to which reasoners with higher and lower cognitive capacities differ regarding this capability which has been called logical intuition (De Neys, 2012).

In two experiments, we used implicit tests of reasoning across several reasoning problems as well as different individual differences measures to investigate both the existence of intuitive logic and individual differences in this capability. In Experiment 1, participants were asked to rate how much they liked the conclusion of several reasoning problems, and results showed that the conclusions to valid statements were rated as more likable than those to invalid statements. The pattern of the data was compatible with the results of the validity judgment task in which participants were explicitly asked to evaluate the logical status of several problems. In Experiment 2, we asked participants to rate the physical brightness of the final conclusion of several arguments. Although participants showed good accuracy in distinguishing high and low contrasts statements, as it is reflected in the significant main effect of brightness, they also judged the conclusion to valid problems to be brighter than invalid ones. To ensure that only the logical intuition rendered the effects, several considerations were implemented. First, a randomization method was used to neutralize the effect of the content on judgment. Moreover, no indication of logical reasoning was mentioned in the instructions.

The logic effect in both liking and brightness tasks is consistent with recent empirical findings and theoretical conceptualizations. Morsanyi and Handley (2012a) found that reasoners liked and endorsed valid problems more than invalid reasoning problems. However, these two kinds of judgments, according to the authors, have distinct underlying mechanisms. Validity judgments, and not liking judgments, were affected by instruction, the figure of the problems, and participants' working memory. On the other hand, manipulating the affective states of the task only affected subjects' liking ratings. One problem with the second and fourth experiments of the Morsanyi and Handley study, however, was that the content was not randomly allocated across conditions. When the contents of problems were randomly assigned to different validity conditions, the effect of logic did not remain significant for complex reasoning problems (Klauer & Singmann, 2013; Singmann et al., 2014). Nevertheless, in a later study with full randomized design, the logic effect was found to be reliable for liking and brightness tasks for simple reasoning problems. In their study, Trippas et al. (2016) demonstrated that valid problems were judged to be more likable and brighter than invalid ones. As the authors argued, these findings imply that a component of logical reasoning is implicit and intuitive in nature.

According to the fluency-affect intuition model (Topolinski, 2011, 2018), the ease with which a coherent argument was conceived produces a sense of fluency, and this fluency results in a positive affect. Although participants are aware of such an affective state, they have no clue regarding its source. Hence, this vague and ambiguous affect is used as a marker for judgments of liking (Reber & Schwarz, 1999), truth (Topolinski & Reber, 2010), and logical validity (Nakamura & Kawaguchi, 2016). In the same vein, the conceptual fluency of coherent valid and believable arguments can produce a positive affect, which in turn may be misattributed to a judgment of logical validity.

Recent dual process theorists have also incorporated the existence of intuitive logic into their models of human reasoning (Bago & De Neys, 2019b; Evans, 2017; Handley & Trippas, 2015; Newman et al., 2017; Pennycook et al., 2015; Stanovich, 2018). The PPT model (Trippas & Handley, 2017) argues that there is not a one to one correspondence between problem features (e.g., believability or validity) and cognitive processing types (i.e., T1 and T2 thinking). This account claims that, depending on the complexity of the task at hand, T1 processes have access to the underlying logical structure of the problem and interestingly, on simple reasoning problems, the logical structure of a statement is more accessible than other features like its believability (Trippas et al., 2017). The fast logic hypothesis of this model is consistent with the observation of higher accuracy and lower speed for logic judgments (Handley et al., 2011; Pennycook, et al., 2014) and the relative immunity of logic judgments from cognitive loading compared to belief judgments (Howarth et al., 2016; Howarth et al., 2018). This fast logic is claimed to be an automatic and intuitive type of thinking which is produced upon the presentation of the problem as one of the many possible initial responses. In this regard, the PPT model is compatible with the three-stage model of analytical engagement (Pennycook et al., 2015) and the logical intuition model (De Neys & Bago, 2017a), according to which, at the onset of the reasoning process, several T1 responses are produced which can be overridden by T2 thinking in later stages of reasoning.

Bago and De Neys (2019b) have argued that one of these responses is a logical intuition and the absolute and relative strength of this response determines the final output of the reasoning process. In a study conducted by Bago and De Neys (2018), it has been shown that in cases in which participants ended with a correct response, this response is mainly produced via an automatic T1 thinking. Moreover, Bago et al. (2019a) found that reasoners have a medium-level knowledge of conflict between their incorrect response and the normative one as is reflected by selecting a second guess closer to the correct answer. This knowledge of conflict indicates that reasoners, even when they come up with a biased response, have a implicit sensitivity to logical rules. Finally, Bago and De Neys (2019b) manipulated the strength of logical intuition by changing the statistical information and found that this manipulation affected the proportion of normative responses selected in a condition favourable for T1 thinking. In a similar vein, Newman et al. (2017) revealed that participants show an early sensitivity to the logic and probability rules as they incorporate these pieces of information into their judgment even when they are given a limited time. The early sensitivity to logic has been supported by electrophysiological evidence, especially by the early activation patterns of components which have attributed to the monitoring (Bence Bago et al., 2018; Banks & Hope, 2014). Altogether, these findings along with the current study results emphasize the need to revise the generic dual process theories regarding the characteristics of each processing type and their interaction. This need has been felt recently as is reflected in the recent revision of the DPT (Evans, 2017, 2018; Stanovich, 2018) and the development of the dual process theory 2.0 (De Neys, 2017).

Another aim of the current study was to investigate individual differences in intuitive logic. As the results showed, participants with higher CA, compared to the low CA participants, rated the conclusion of valid arguments more likable than those of invalid ones. The individual difference findings in the liking judgment task were replicated in Experiment 2 and were similar to the pattern of the findings in the validity judgment task of Experiment 1. However, controlling the explicit reasoning measure rendered the predictive power of CA non-significant. These results indicate that individual differences in liking judgments arise primarily from engagement in deliberative thinking. The logic effect in the brightness task, on the other hand, was not different among participants with different CA. In other words, smarter reasoners had no advantages in the implicit sensitivity to logic compared to their low CA counterparts, and if anything, lower CA participants were better at rating valid and invalid statements with different brightness ratings when other variables were controlled.

One hypothesis could account for the lack of relationship between cognitive measures and the intuitive judgments. According to this hypothesis, which we call the continuum hypothesis of analytical reasoning engagement, analytical reasoning engagement varies on a continuum (See Figure 4) with the most possible engagement of analytical thinking (i.e., explicit reasoning) on the one side and the least possible engagement of such thinking (i.e., implicit reasoning) on the other side. According to this hypothesis, validity judgment is an explicit reasoning task, and liking and brightness judgments are more implicit measures with brightness judgment at the furthest end of the continuum. In the liking judgment task, there are no external criteria besides the logical and believability status of the problem, and it is probable that many participants rely on these features in their judgment. As the relationships between the logic effect in liking and endorsement judgment shows, liking judgments measure both explicit and implicit sensitivity to logic. However, providing the physical contrast as an external criterion in the brightness rating task gives participants a clear measure for their judgment. Participants are instructed to rate the physical brightness of some statements which varies in their level of contrast, and there is no apparent reason for relying on the logical status of the problems in doing so. As Trippas et al. (2016) explained: "[t]o the layman, the criteria for brightness are more narrowly defined than for liking and do not plausibly include the semantic or syntactic content of the materials." This hypothesis considers brightness judgment as a purer measure of logical intuition and argues that such pure measures of intuition render individual differences in cognitive ability as less significant.

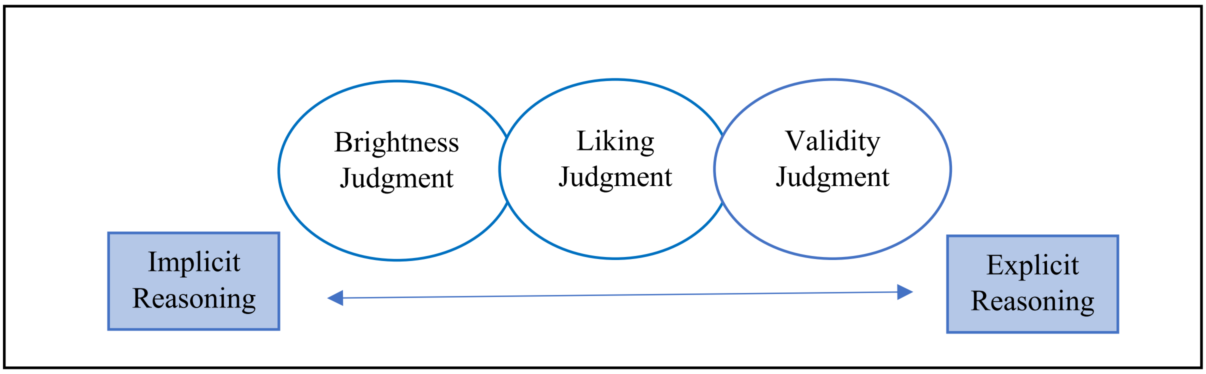

Figure 4 An overview of the continuum hypothesis of analytical engagement

One critical test for the continuum hypothesis is the use of the time pressure paradigm. The continuum hypothesis predicts that limiting the response time should not affect the logic effect in brightness judgments since this effect is primarily intuitive. However, the logic effect in liking judgments should be decreased in the speeded task, because time pressure would hinder the explicit reasoning component of liking judgments. More importantly, this hypothesis predicts the lack of logic by ability interactions in both brightness and liking judgment tasks, which implies that purifying intuitive measures from explicit reasoning removes the relationship between individual differences in cognitive abilities and intuitive logic.

In general, the findings imply that individual differences in logical intuition, and more generally, in T1 thinking, are unrelated to cognitive abilities. In other words, participants with different cognitive ability and cognitive styles diverge, not at the beginning of the reasoning process, rather at the later point of this process (Betsch & Glöckner, 2010; Darlow & Sloman, 2010; De Neys & Bonnefon, 2013; Morsanyi & Handley, 2012a; Stanovich & West, 2008b; Toplak et al., 2014). However, it should be noted that the current study investigated individual differences in respect to cognitive ability and styles. It could be the case that reasoners show an early divergence in respect to other cognitive factors, such as the quantity of their attention to reasoning problems (Mata et al., 2017), the efficacy of their automatized logical rule (Nakamura & Kawaguchi, 2016; Stanovich, 2018), their metacognitive awareness of normative responses (Thompson & Johnson, 2014), or their level of conflict detection specificity (Bago & De Neys, 2019a ; Frey et al., 2017).

Although these findings are not necessarily incompatible with the early divergence accounts, they are inconsistent with several studies which have shown individual differences in cognitive abilities is related to in intuitive logic. As we reviewed earlier, Thompson et al. (2018) found that, when participants were instructed to judge based on the logic, those with higher cognitive abilities (CA & ACS) had higher accuracy and lower response latency. The authors concluded that considering the underlying logical structure is the default response of reasoners with higher abilities and these participants are better at this capability. Moreover, Thompson and Johnson (2014) demonstrated that even when participants were asked to respond as quickly as possible, those with higher CA and ACS are doing better to respond based on the logical rules. Besides using different measures of logical intuition (with varying levels of complexity) and cognitive abilities (such as the use of self-report measures of cognitive style by Thompson and colleagues), there are several explanations that could account for the anomaly. First, Thompson and colleagues found that logic judgments can be accomplished more accurately than belief judgments in the case of base-rate and syllogistic reasoning problems. However, one can argue that these judgments were made by possible engagement of T2 thinking. Researchers tried to address this issue in their second experiment by applying a time pressure of 20 seconds, however, this amount of time still leaves the door open for engaging in deliberative thinking. As Bago and De Neys (2017a) found, the reading time for the long and short version of base-rate problems are approximately 12 seconds and 3 seconds, respectively, which are significantly lower than the amount of the time that participants had in Thompson et al.'s study. Second, it could be possible that the better performance of reasoners with higher CA reflects using superficial strategies (e.g., atmosphere) rather than assessing logic. As Handley et al. (2011) demonstrated, presenting the cue (Judgment based on logic or belief) prior to the conclusion presentation, which was done in Thompson et al.'s study, may encourage participants to find some strategic shortcuts for responding.

The current study aimed to investigate the existence of sensitivity to logic and individual differences in this sensitivity. An important question is what are the underlying mechanisms that bring such intuitive judgments about? We did not explicitly address the mechanisms of logical intuition; however, we can speculate on some possible mechanisms. Several theories argue that the coherency of information is among one of the mechanisms that logical intuition relies on (Betsch & Glöckner, 2010; Topolinski, 2011). For example, as Trippas et al. (2016) described, the conceptual fluency that arises from coherent information can produce a positive affect. This affect can be used as a cue to a judgment of logical validity. One supporting evidence of considering the underlying coherence of information is the conflict detection findings of the current study. Similar to conflict detection studies, in the current study, logical and believability status of reasoning problems may either conflict or not. In the brightness judgment task, participants judged no-conflict problems to be significantly brighter than conflict problems ($M= 3.92$, $SD= .60$ for no-conflict problems and $M= 3.83$, $SD= .65$ for conflict problems), $t(46) = 2.258$, $p= .029$, $d= .33$. Similarly, in the liking judgment tasks of Experiment 1 and 2 combined, we found a significant conflict sensitivity, $t(93) = 3.17$, $p= .002$, $d= .32$.

Considering the coherency of information as a cue for judgment implies that reasoners are aware of the logical structure of the problem intuitively. This structure detection may be similar to the knowledge storage in the mindware continuum hypothesis (Stanovich, 2018). According to this model, the reasoning process comprises three parameters of mindware, detection, and inhibition, and reasoners have similar mindware and detection capability and they diverge only at the inhibition stage (Stanovich & West, 2008b). Our findings are consistent with this model in respect to the similar mindware for high and low ability participants. Thus, the ease with which one can detect the structure (i.e., mindware) is a possible mechanism of logical intuition which is not different for participants in respect to their CA scores (Frey et al., 2017). One supporting piece of evidence for this hypothesis comes from the regression analysis according to which the logic index in validity judgment predicted the liking logic index. We need, however, to be cautious about this finding since the validity judgment task of the first experiment may not be a suitable measure of mindware since validity and believability of these problems were manipulated orthogonally. Future research could use abstract problems as a better index of mindware.

The awareness of the underlying logical structure reveals another proposed mechanism of logical intuition. According to De Neys and Bago (2017a), at least one heuristic intuitive and one logical intuitive response is triggered upon the presentation of a reasoning problem. The final output of the reasoning process depends on the relative and absolute strength or activation level of these two responses. To test this account, Bago and De Neys (2019b) manipulated the strength of logical and heuristic intuition by changing the statistical or stereotypical information in the base-rate task. They found that not only the absolute strength of two intuitions determined the proportion of normative or non-normative responses but also the relative strength of those intuitions affected the conflict detection rates. An interesting research question is whether manipulating the strength of logical intuition can influence the liking or brightness ratings of participants as other intuitive measures? One way to manipulate the strength of logical intuition in the liking and brightness judgment tasks is to use indeterminate problems. All of the invalid problems in our study were determinately invalid. In other words, no model can be found in which the problem could be logically valid. On the other hand, for indeterminate invalid problems, it is possible to come up with a model in which the conclusion is consistent with the premises. An example of such a problem which was taken from Evans et al. (1999) is as follows:

All artists are beekeepers,
Lisa is a beekeeper,
Therefore, Lisa is an artist

By applying indeterminate problems, we can reduce the strength of logical intuition. If participants are sensitive to the logical status of the problem, they should judge valid determinate problems as more likable and brighter than indeterminate ones. This method originally was proposed by Trippas et al.'s (2016) as a suitable paradigm to reveal that whether creating a coherent semantic model or coherent logical model is the effective factor that creates fluency.

Future Directions

Since logical intuition is a relatively new concept in the realm of judgment and reasoning, there are few studies focusing on this capability and its cognitive mechanisms. Although several models in the dual process theory incorporate this concept as one of the main components of the reasoning process, more research is needed for a better understanding of logical intuition, and more generally, how reasoning and intuition interact. In the previous section, we proposed some possible studies to further explore logical intuition and replicate the current findings with different methods. For example, applying a time pressure paradigm could provide us with a pure measure of logical intuition and using indeterminate problems and abstract problems could broaden our knowledge of intuitive logic and its mechanisms.

Since investigating the conditions under which an unexpected effect disappears is a characteristic of the development of a program, as Kahneman and Frederick described (2002, representativeness), delineating the boundary conditions of logical intuition could be the next step in studying this phenomenon. Several studies have already addressed the boundary condition of logical intuition in term of arguments' complexity (Brisson et al., 2018; Frey et al., 2017; Klauer & Singmann, 2013; Singmann et al., 2014). Individual differences also can provide a suitable opportunity to investigate the boundary conditions (Pennycook et al., 2012). By making a connection between individual differences studies and studies on the boundary condition of logical intuition, we could investigate the individual differences in the different level of complexity. For example, Klauer and Singmann (2013) found that reasoners have no intuitive sensitivity toward logic on more complex reasoning problem. However, it may be possible to find the logic effect on complex problems for high ability reasoners. Individual differences studies could yield more precise picture of human reasoning abilities.

Conclusion

In two studies, we found that reasoners are intuitively sensitive to the logical validity of a reasoning problem as they judged valid problems to be more likable and brighter than their counterparts. More importantly, the results revealed that reasoners with high and low cognitive abilities showed similar sensitivity to logic. These findings contradict the traditional dual process accounts regarding the characteristics of each type of thinking, while support the late divergence models of reasoning in respect to the individual differences in cognitive abilities. In other words, the results emphasise the necessity for major revision in the classic views of human cognition which have been suggested recently under the umbrella of the dual process theories 2.0.

References

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

Bago, Bence, & De Neys, W. (2017b). Rise and fall of conflicting intuitions during reasoning. In *CogSci*.

Bago, B., & De Neys, W. (2019b). Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view. *Thinking & Reasoning*, 1-30

Bago, B., & De Neys, W. (2018). The smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking and Reasoning*.

Bago, Bence, Frey, D., Vidal, J., Houdé, O., Borst, G., & De Neys, W. (2018). Fast and slow thinking: Electrophysiological evidence for early conflict sensitivity. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.07.017

Bago, B., Raoelison, M., & De Neys, W. (2019a). Second-guess: Testing the specificity of error detection in the bat-and-ball problem. Acta psychologica, 193, 214-228.

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*. https://doi.org/10.1027/1618-3169.53.1.77

Ball, L., Thompson, V., & Stupple, E. (2017). Conflict and dual process theory: The case of belief bias. Retrieved from https://derby.openrepository.com/derby/handle/10545/622577

Banks, A. P., & Hope, C. (2014). Heuristic and analytic processes in reasoning: An event-related potential study of belief bias. *Psychophysiology*. https://doi.org/10.1111/psyp.12169

Betsch, T., & Glöckner, A. (2010). Intuition in judgment and decision making: Extensive thinking without effort. *Psychological Inquiry*. https://doi.org/10.1080/1047840X.2010.517737

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory and Cognition*, *38*(2), 186–196. https://doi.org/10.3758/MC.38.2.186

Brisson, J., Schaeken, W., Markovits, H., & De Neys, W. (2018). Conflict Detection and Logical Complexity. *Psychologica Belgica*, *58*(1), 318. https://doi.org/10.5334/pb.448

Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*. https://doi.org/10.1002/wcs.34

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

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

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking and Reasoning*, *20*(2), 169–187. https://doi.org/10.1080/13546783.2013.854725

De Neys, W. (2015). *Heuristic bias and conflict detection during thinking*. *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 62). Elsevier Ltd. https://doi.org/10.1016/bs.plm.2014.09.001

De Neys, W. (2017). *Dual process theory 2.0*. *Dual Process Theory 2.0*. https://doi.org/10.4324/9781315204550

De Neys, W. (2017). Bias, conflict, and fast logic: towards a hybrid dual process future?. In *Dual Process Theory 2.0* (pp. 55-73). Routledge.

De Neys, W., & Bonnefon, J.-F. (2013). The Whens and Whys of Individual Differences in individual thinking biases. *Trends in Cognitive Sciences*, *17*(4), 172–178. https://doi.org/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). https://doi.org/10.1371/journal.pone.0015954

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, *113*(1), 45–61. https://doi.org/10.1016/j.cognition.2009.07.009

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

De Neys, W., Moyens, E., & Ansteenwegen, D. V. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective and Behavioral Neuroscience*, *10*(2), 208–216. https://doi.org/10.3758/CABN.10.2.208

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-013-0384-5

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2008.02113.x

Dhar, R., & Gorlin, M. (2013). A dual-system framework to understand preference construction processes in choice. *Journal of Consumer Psychology*, *23*(4), 528–542. https://doi.org/10.1016/j.jcps.2013.02.002

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*. https://doi.org/10.1037/0003-066X.49.8.709

Evans, Jonathan St. B. T., Handley, S., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, *35*(7), 1772–1784. https://doi.org/10.3758/BF03193509

Evans, Jonathan St. B. T. , Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Evans, Jonathan St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, Jonathan St B.T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2003.08.012

Evans, Jonathan St B.T. (2017). Dual process theory: perspectives and problems. In Dual Process Theory 2.0 (pp. 145-164). Routledge.

Evans, Jonathan St B.T. (2018). Dual-process theories. In International Handbook of Thinking and Reasoning (pp. 151-165). Routledge.

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

Evans, Jonathan St B.T., Handley, S. J., & Bacon, A. M. (2009). Reasoning under time pressure: a study of causal conditional inference. *Experimental Psychology*, *56*(2), 77–83.

https://doi.org/10.1027/1618-3169.56.2.77

Evans, Jonathan St B.T., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *54*(3), 935–958. https://doi.org/10.1080/713755983

Evans, Jonathan St B.T.., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1495–1513. https://doi.org/10.1037/0278-7393.25.6.1495

Evans, Jonathan St B.T., Handley, S. J., Neilens, H., & Over, D. (2008). Understanding causal conditionals: A study of individual differences. *Quarterly Journal of Experimental Psychology*, *61*(9), 1291–1297. https://doi.org/10.1080/17470210802027961

Evans, Jonathan St B.T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691612460685

Evans, Jonathan St B T. (2016). Intuition and Reasoning : A Dual-Process Perspective Intuition and Reasoning : A Dual-Process Perspective, *7965*(September). https://doi.org/10.1080/1047840X.2010.521057

Frankish, K., & Evans, J. S. B. T. (2012). The duality of mind: An historical perspective. In *In Two Minds: Dual Processes and Beyond*. https://doi.org/10.1093/acprof:oso/9780199230167.003.0001

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. https://doi.org/10.1257/089533005775196732

Frey, D., Johnson, E. D., & De Neys, W. (2017). Individual differences in conflict detection during reasoning. *The Quarterly Journal of Experimental Psychology*, 1–52. https://doi.org/10.1080/17470218.2017.1313283

Gigerenzer, G., & Gaissmaier, W. (2010). *Heuristic Decision Making*. *SSRN*. https://doi.org/10.1146/annurev-psych-120709-145346

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*. https://doi.org/10.1016/j.actpsy.2012.01.009

Goel, V., & Dolan, R. J. (2003). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2003.07.027

Gould, S. J. (1989). The Streak of Streaks. *CHANCE*, *2*(2), 10–16. https://doi.org/10.1080/09332480.1989.10554932

Hagmayer, Y., & Sloman, S. A. (2009). Decision Makers Conceive of Their Choices as Interventions, *138*(1), 22–38. https://doi.org/10.1037/a0014585

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, Beliefs, and Instruction: A Test of the Default Interventionist Account of Belief Bias. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(1), 28–43. https://doi.org/10.1037/a0021098

Handley, S. J., & Trippas, D. (2015). *Dual processes and the interplay between knowledge and structure: A new parallel processing model. Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 62). Elsevier Ltd. https://doi.org/10.1016/bs.plm.2014.09.002

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*(9), 1448–1457. https://doi.org/10.1037/xlm0000248

Heim, A. (1970). AH 4 Group Test of General Intelligence. Manual. By A (lice) W (inifred) Heim. Rev. Ed. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=AH4+Group+Test+of+General+Intelligence+Heim%2C+1970&btnG=

Howarth, S. & Handley, S. (2016). Belief bias, base rates and moral judgment: re-evaluating the default interventionist dual process account. In *The Thinking Mind* (pp. 97-111). Psychology Press.

Howarth, S., Handley, S. J., & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief–logic conflict. *Memory and Cognition*, *44*(2). https://doi.org/10.3758/s13421-015-0555-x

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

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*. https://doi.org/10.1016/j.jcps.2014.12.002

Kahneman, D., & Frederick, S. (2002). Heuristics of Intuitive Judgment: Extensions and Applications. In *Heuristics of Intuitive Judgment: Extensions and Applications*. https://doi.org/10.1038/2251090a0

Kahneman, Daniel, & Frederick, S. (2005). A model of Heuristic Judgement. In *The Cambridge handbook of thinking and reasoning*. https://doi.org/10.1111/cogs.12119

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*. https://doi.org/10.1111/1467-8624.00232

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852–884. https://doi.org/10.1037/0033-295X.107.4.852

Klauer, K. C., & Singmann, H. (2013). Does logic feel good? Testing for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(4), 1265–1273. https://doi.org/10.1037/a0030530

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/S0022-0965(02)00121-2

Lea, R. B. (1995). On-Line Evidence for Elaborative Logical Inferences in Text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/0278-7393.21.6.1469

Lea, R. B., O'Brien, D. P., Fisch, S. M., Noveck, I. A., & Braine, M. D. S. (1990). Predicting propositional logic inferences in text comprehension. *Journal of Memory and Language*, *29*(3), 361–387. https://doi.org/10.1016/0749-596X(90)90005-K

Leo, P. D., & Greene, A. J. (2008). Is awareness necessary for true inference? *Memory and Cognition*. https://doi.org/10.3758/MC.36.6.1079

Mata, A., Ferreira, M. B., Voss, A., & Kollei, T. (2017). Seeing the conflict: an attentional account of reasoning errors. *Psychonomic Bulletin and Review*, *24*(6), 1980–1986. https://doi.org/10.3758/s13423-017-1234-7

McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, Type II errors, and false-positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*. https://doi.org/10.1016/j.jcps.2015.05.006

Melnikoff, D. E., & Bargh, J. A. (2018). The Mythical Number Two. *Trends in Cognitive Sciences*, *22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Mevel, K., Poirel, N., Rossi, S., Cassotti, M., Simon, G., Houdé, O., & De Neys, W. (2015). Bias detection: Response confidence evidence for conflict sensitivity in the ratio bias task. *Journal of Cognitive Psychology*, *27*(2), 227–237. https://doi.org/10.1080/20445911.2014.986487

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2010.07.004

Morsanyi, K, & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, *99*(1), 18–36. https://doi.org/10.1016/j.jecp.2007.08.003

Morsanyi, K, & Handley, S. J. (2012a). Logic feels so good-I like it! evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(3), 596–616. https://doi.org/10.1037/a0026099

Morsanyi, K, & Handley, S. (2012b). Does thinking make you biased? The case of the engineers and lawyer problem. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*, 2049–2054.

Nakamura, H., & Kawaguchi, J. (2016). People like logical truth: Testing the intuitive detection of logical value in basic propositions. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0169166

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1154–1170. https://doi.org/10.1037/xlm0000372

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*. https://doi.org/10.1080/02724980343000116

Newstead, S. E., Pollard, P., Evans, J. S. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*(3), 257–284. https://doi.org/10.1016/0010-0277(92)90019-E

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-01193-y

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*. https://doi.org/10.3389/neuro.11.010.2008

Peirce, J. W. (2009) Generating stimuli for neuroscience using PsychoPy. Front. Neuroinform. 2:10. doi:10.3389/neuro.11.010.2008

Pennycook, G. (2017). A perspective on the theoretical foundation of dual process models. In *Dual Process Theory 2.0*. https://doi.org/10.4324/9781315204550

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition*, *42*(1), 1–10. https://doi.org/10.3758/s13421-013-0340-7

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning? *Cognition*, *124*(1), 101–106. https://doi.org/10.1016/j.cognition.2012.04.004

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72.

https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., Neys, W. De, Evans, J. S. B. T., Stanovich, K. E., & Thompson, V. A. (2018). The Mythical Dual-Process Typology. *Trends in Cognitive Sciences*, *22*(8), 667–668. https://doi.org/10.1016/j.tics.2018.04.008

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(2), 544–554. https://doi.org/10.1037/a0034887

Qualtrics, I. (2013). Qualtrics. *Provo, UT, USA*.

Rader, A. W., & Sloutsky, V. M. (2002). Processing of Logically Valid and Logically Invalid Conditional Inferences in Discourse Comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.28.1.59

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgements of truth. *Consciousness and Cognition*, *8*(1), 338–342.

Reber, R., & Unkelbach, C. (2010). The Epistemic Status of Processing Fluency as Source for Judgments of Truth. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-010-0039-7

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

Rucker, D. D., McShane, B. B., & Preacher, K. J. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology*. https://doi.org/10.1016/j.jcps.2015.04.004

Singmann, H., Klauer, K. C., & Kellen, D. (2014). Intuitive logic revisited: New data and a bayesian mixed model meta-analysis. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094223

Sloman, S. (2014). Two systems of reasoning: An update. Retrieved from http://psycnet.apa.org/record/2014-08812-005

Sloman, Steven A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.119.1.3

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Retrieved from https://www.taylorfrancis.com/books/9781135687564

Stanovich, K. E. (2018). Miserliness in human cognition: the interaction of detection, override and mindware. *Thinking & Reasoning*. https://doi.org/10.1080/13546783.2018.1459314

Stanovich, K. E., & West, R. F. (1998). Individual Differences in Rational Thought. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.127.2.161

Stanovich, K. E., & West, R. F. (2008). On the Relative Independence of Thinking Biases and Cognitive Ability. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.94.4.672

Stupple, E. J. N., Ball, L. J., Evans, J. S. B. T., & Kamal-, E. (2011). When logic and belief collide : Individual differences in reasoning times support a selective processing model When logic and belief collide : Individual differences in reasoning times support a selective processing model, *23*(April 2015), 37–41. https://doi.org/10.1080/20445911.2011.589381

Thompson, V. A. (2014). *What intuitions are... and are not*. *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 60). https://doi.org/10.1016/B978-0-12-800090-8.00002-0

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

Thompson, V. A., & Newman, I. R. (2017). Logical intuitions and other conundra for dual process theories. In *Dual Process Theory 2.0*. https://doi.org/10.4324/9781315204550

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. S. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology: General*, *147*(7), 945–961. https://doi.org/10.1037/xge0000457

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*. https://doi.org/10.1016/j.cognition.2012.09.012

Thompson, V., & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind and Society*, *11*(1), 93–105. https://doi.org/10.1007/s11299-012-0100-6

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking and Reasoning*. https://doi.org/10.1080/13546783.2013.844729

Topolinski, S. (2011). A process model of intuition. *European Review of Social Psychology*, *22*(1), 274–315. https://doi.org/10.1080/10463283.2011.640078

Topolinski, S. (2018). The sense of coherence: How intuition guides reasoning and thinking. In International Handbook of Thinking and Reasoning (pp. 559-574). Routledge.

Topolinski, S., Likowski, K. U., Weyers, P., & Strack, F. (2009). The face of fluency: Semantic coherence automatically elicits a specific pattern of facial muscle reactions. *Cognition and Emotion*. https://doi.org/10.1080/02699930801994112

Topolinski, S., & Reber, R. (2010). Immediate truth - Temporal contiguity between a cognitive problem and its solution determines experienced veracity of the solution. *Cognition*, *114*(1), 117–122. https://doi.org/10.1016/j.cognition.2009.09.009

Topolinski, S., & Strack, F. (2009). The analysis of intuition: Processing fluency and affect in judgements of semantic coherence. *Cognition and Emotion*. https://doi.org/10.1080/02699930802420745

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

Trippas, D., & Handley, S. J. (2017). The parallel processing model of belief bias: review and extensions. In *Dual Process Theory 2.0* (pp. 28-46). Routledge.

Trippas, D., Handley, S. J., & Verde, M. F. (2013). The SDT model of belief bias: Complexity, time, and cognitive ability mediate the effects of believability. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/a0032398

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*. https://doi.org/10.1037/xlm0000248

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking and Reasoning*, *21*(4), 431–445. https://doi.org/10.1080/13546783.2015.1016450

Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory and Cognition*, *45*(4), 539–552. https://doi.org/10.3758/s13421-016-0680-1

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*. https://doi.org/10.1037/0033-295X.90.4.293

Winkielman, P., Schwarz, N., Fazendeiro, T. a, & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. *The Psychology of Evaluation: Affective Processes in Cognition and Emotion.*, 189–217. https://doi.org/http://dx.doi.org/10.4324/9781410606853

Appendices

Appendix A: Materials

Appendix 1 Modus ponens and modus tollens conditionals used in Experiment 1 and 2

Nr	Conditional
1	If Aden is [sober/drunk] then he consumed a [tiny/large] amount of alcohol
2	if an airplane [collides with/goes over] a mountain then it will [crash/fly]
3*	If Andrew consistently eats [junk/healthy] food, then he is [fat/lean]
4	If a bear [catches/fails to catch] a fish it is [quick/slow]
5	If a child is [crying/laughing] then it is [sad/happy]
6	If a child is [clean/dirty] then it has been playing in the [bath/mud]
7*	If chips are [frozen/cooked] then they taste [disgusting/delicious]
8*	If a clock is [broken/working] then its hands [stand still/move]
9	If a new computer is [high/low] end then it is [expensive/cheap]
10*	If a dog [is barking/has its mouth shut] then it is [loud/quiet]
11	If Jacky is [hungry/full] then she has eaten [too little/too much]
12	If John is [in/out of] the water then he is [wet/dry]
13	If a racehorse is [slow/fast] then it will [lose/win] many races
14	If a cup of tea contains [sugar/salt] then it tastes [sweet/salty]
15	If there is a [strong/weak] wind then the kite will [go up in the air/stay on the ground]
16	If it is [winter/summer] then it is [freezing/hot] outside

*Used only in the second experiment

Appendix 2 Affirmation and denial disjunctions used in the experiment 1 and 2

Nr	Disjunction
1	Either beers are drinks or they are boats
2	Either cats are mammals or they are plants
3	Either circles are round or they are square
4	Either flowers are organic or they are vehicles
5	Either frogs are amphibians or they are birds
6	Either hammers are tools or they are vegetables
7*	Either Labradors are dogs or they are fish
8	Either mice eat cheese or they eat steel
9*	Either monkeys are primates or they are rodents
10*	Either obese people are fat or they are skinny
11*	Either robbers are criminals or they are drinks
12	Either roses are flowers or they are machines
13	Either sentences are made out of words or bricks
14	Either the sky is blue or it is green
15	Either skyscrapers are huge or they are tiny
16	Either snakes are reptiles or they are trees

*Used only in the second experiment

Appendix 3 Syllogistic figures used in the Experiment 1 and 2

Nr	Believability	Validity	Structure	Conclusion
1	Believable	Valid	AE1_E1	No A are C
2	Believable	Valid	AE3_E1	No A are C
3	Believable	Valid	EA2_E1	No A are C
4*	Believable	Valid	AI4_I1	Some A are C
5*	Believable	Valid	AE3_E2	No C are A
6	Believable	Valid	IA1_I2	Some C are A
7	Believable	Valid	AI4_I2	Some C are A
8	Believable	Valid	AI2_I2	Some C are A
9	Believable	Invalid	AE1_I1	Some A are C
10	Believable	Invalid	AE3_I1	Some A are C
11	Believable	Invalid	EA2_I1	Some A are C
12*	Believable	Invalid	AI4_E1	No A are C
13*	Believable	Invalid	AE3_I2	Some C are A
14	Believable	Invalid	IA1_E2	No C are A
15	Believable	Invalid	AI4_E2	No C are A
16	Believable	Invalid	AI2_E2	No C are A
17*	Unbelievable	Valid	AE3_E2	No C are A
18	Unbelievable	Valid	IA1_I2	Some C are A
19	Unbelievable	Valid	AI4_I2	Some C are A
20	Unbelievable	Valid	AI2_I2	Some C are A
21	Unbelievable	Valid	AE1_E1	No A are C
22	Unbelievable	Valid	AE3_E1	No A are C
23	Unbelievable	Valid	EA2_E1	No A are C
24*	Unbelievable	Valid	AI4_I1	Some A are C
25*	Unbelievable	Invalid	AE3_I2	Some C are A
26	Unbelievable	Invalid	IA1_E2	No C are A
27	Unbelievable	Invalid	AI4_E2	No C are A
28	Unbelievable	Invalid	AI2_E2	No C are A
29	Unbelievable	Invalid	AE1_I1	Some A are C
30	Unbelievable	Invalid	AE3_I1	Some A are C
31	Unbelievable	Invalid	EA2_I1	Some A are C
32*	Unbelievable	Invalid	AI4_E1	No A are C

In the structure column of the table above, the first letter represents the quantifier of the major premise, the second one represents the quantifier of the minor premise, and the last letter represents the quantifier of the final conclusion. A (all) = universal affirmative, I (some) = particular affirmative, E (no) = universal negative. The first number (1-4) demonstrates the Figure of the syllogism (1 = AB-BC, 2 = BA-CB, 3 = AB-CB, 4 = BA-BC), and the second number demonstrates the conclusion direction (1 = A-C, 2 = C-A).

*Used only in the second experiment

Appendix 4 Category-member pairs for the first experiment

Category	Member
marsupials	kangaroos
birds	parrots
boats	canoes
cars	Volvos
criminals	robbers
furniture	desks
dogs	labradors
drinks	beers
fish	salmons
fruits	bananas
insects	bees
reptiles	lizards
tools	hammers
trees	eucalyptus
vegetables	cabbages
weapons	swords
clothes	jeans
musics	hip-hop
buildings	churches
meats	porks
currencies	dollars
organs	lungs
movies	comedies
cats	tabby

Appendix 5 Category-member pairs for the second experiments

Category	Member 1	Member 2
weapons	guns	swords
birds	parrots	sparrows
boats	kayaks	canoes
cars	BMWs	Volvos
currencies	dollars	euros
furniture	desks	sofas
dogs	spaniels	labradors
drinks	beers	whiskeys
fish	trouts	salmons
fruits	bananas	peaches
insects	bees	beetles
reptiles	lizards	crocodiles
tools	hammers	saws
trees	eucalyptus	willows
vegetables	carrots	cabbages
criminals	robbers	murderers

1. In a study, 1000 people were tested. Among the participants, there were 4 men and 996 women. Jo is a randomly chosen participant of this study. Jo is 23 years old and is finishing a degree in engineering. On Friday nights, Jo likes to go out cruising with friends while listening to loud music and drinking beer. What is most likely?
a. Jo is a man
b. Jo is a woman

2. In a study, 1000 people were tested. Among the participants, there were 5 engineers and 995 lawyers. Jack is a randomly chosen participant of this study. Jack is 36 years old. He is not married and is somewhat introverted. He likes to spend his free time reading science fiction and writing computer programs. What is most likely?
a. Jack is an engineer
b. Jack is a lawyer

3. In a study, 1000 people were tested. Among the participants, there were three who live in a condo and 997 who live in a farmhouse. Kurt is a randomly chosen participant of this study. Kurt works on Wall Street and is single. He works long hours and wears Armani suits to work. He likes wearing shades. What is most likely?
a. Kurt lives in a condo
b. Kurt lives in a farmhouse

4. In a study, 1000 people were tested. Among the participants, there were 997 nurses and 3 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

5. In a study, 1000 people were tested. Among the participants, there were four whose favorite series is Star Trek and 996 whose favorite series is Days of Our Lives. Jeremy is a randomly chosen participant of this study. Jeremy is 26 and is doing graduate studies in physics. He stays at home most of the time and likes to play video-games. What is most likely?
a. Jeremey's favorite series is Star Trek
b. Jeremey's favorite series is Days of Our Lives

6. In a study, 1000 people were tested. Among the participants, there were 5 sixteen-year-olds and 995 fifty-year-olds. Ellen is a randomly chosen participant of this study Ellen likes to listen to hip hop and rap music. She enjoys wearing tight shirts and jeans. She's fond of dancing and has a small nose piercing. What is most likely?
a. Ellen is sixteen
b. Ellen is fifty

7. 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]

8. 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]

9. 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]

10. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? _____ days [correct answer = 4 days; intuitive answer = 9]

11. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? ______ students [correct answer = 29 students; intuitive answer = 30]

12. A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? _____ dollars [correct answer = $20; intuitive answer = $10]

13. Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money [correct answer = c, because the value at this point is $7,000; intuitive response = b].

Appendix B: Statistical Tables

Appendix 7 A summary of the main effects and interactions of ANOVA performed in the Experiment 1 and 2

	Source	F	df	P Value	η^2_p
Liking Judgment (Exp 1)	Logic	50.97	(1, 42)	.000	.548
	Belief	74.58	(1, 42)	.000	.640
	Problem Type	12.92	(2, 84)	.000	.235
	Cognitive Ability	0.96	(1, 42)	.333	.022
	Logic × Belief	6.99	(1, 42)	.011	.143
	Logic × CA	4.17	(1, 42)	.047	.090
	Belief × CA	5.06	(1, 42)	.030	.107
	Logic × Problem Type	5.55	(2, 84)	.005	.117
	Belief × Problem Type	6.70	(1.64,68.98)	.004	.138
Validity Judgment (Exp 1)	Logic	336.6	(1, 45)	.000	.882
	Belief	1.60	(1, 45)	.213	.034
	Problem Type	2.65	(1.76, 79.1)	.084	.056
	Cognitive Ability	.704	(1, 45)	.007	.151
	Logic × Belief	12.18	(1, 45)	.001	.213
	Logic × CA	5.69	(1, 45)	.021	.112
	Belief × CA	.370	(1, 45)	.546	.008
	Logic × Problem Type	1.73	(2, 90)	.183	.037
	Belief × Problem Type	.090	(1.90, 85.55)	.905	.002
Brightness Judgment (Exp 2)	Logic	10.37	(1, 44)	.002	.191
	Belief	16.05	(1, 44)	.000	.267
	Problem Type	3.55	(1.52, 66.73)	.046	.075
	Brightness	25.50	(1, 44)	.000	.367
	Cognitive Ability	4.61	(1, 44)	.037	.095
	Cognitive Style	2.12	(1, 44)	.152	.046
	Logic × Belief	4.71	(1, 44)	.035	.097
	Logic × CA	1.50	(1, 44)	.227	.033
	Belief × CA	1.92	(1, 44)	.173	.042
	Logic × ACS	1.13	(1, 44)	.293	.025
	Belief × ACS	.030	(1, 44)	.864	.001
	Brightness × CA	5.20	(1, 44)	.027	.106

	Logic × Problem Type	3.23	(2,88)	.044	.068
	Brightness × ProbType	3.46	(2,88)	.036	.073
Liking Judgment (Exp 2)	Logic	27.48	(1, 43)	.000	.390
	Belief	77.04	(1, 43)	.000	.642
	Problem Type	23.35	(1.59, 68.35)	.000	.352
	Cognitive Ability	1.90	(1, 43)	.175	.042
	Cognitive Style	2.53	(1, 43)	.119	.055
	Logic × Belief	3.17	(1, 43)	.082	.069
	Logic × CA	4.82	(1, 43)	.034	.101
	Belief × CA	.035	(1, 43)	.853	.001
	Logic × ACS	.190	(1, 43)	.665	.004
	Belief × ACS	.413	(1, 43)	.524	.010
	Logic × Problem Type	5.78	(2, 86)	.004	.118
	Belief × Problem Type	10.14	(1.55, 66.63)	.000	.191

Appendix C: Ethics Letter of Approval

Appendix 8 The ethics approval of the experiments from the Human Science Ethics Sub-Committee

Human Sciences Ethics Sub-Committee
Macquarie University, North Ryde
NSW 2109, Australia

MACQUARIE University

17/09/2018

Dear Professor Handley,

Reference No: 5201832794326
Project ID: 3279
Title: The bright homunculus in our head: The underlying mechanisms of intuitive sensitivity to logical validity

Thank you for submitting the above application for ethical review. The Human Sciences Subcommittee has considered your application.

I am pleased to advise that ethical approval has been granted for this project to be conducted by Professor Simon Handley, and other personnel: M Omidreza Ghasemi, Dr Stephanie Howarth.

This research meets the requirements set out in the National Statement on Ethical Conduct in Human Research 2007, (updated July 2018).

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the National Statement, available from the following website: https://www.nhmrc.gov.au/_files_nhmrc/file/publications/national-statement-2018.pdf.
2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol. You will be sent an automatic reminder email one week from the due date to remind you of your reporting responsibilities.
3. All adverse events, including unforeseen events, which might affect the continued ethical acceptability of the project, must be reported to the subcommittee within 72 hours.
4. All proposed changes to the project and associated documents must be submitted to the subcommittee for review and approval before implementation. Changes can be made via the Human Research Ethics Management System.

The HREC Terms of Reference and Standard Operating Procedures are available from the Research Services website: https://www.mq.edu.au/research/ethics-integrity-and-policies/ethics/human-ethics.

It is the responsibility of the Chief Investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Faculty Ethics Officer.

The Human Sciences Subcommittee wishes you every success in your research.

Yours sincerely,

NSweller

Dr Naomi Sweller

Chair, Human Sciences Ethics Sub-Committee

The Faculty Ethics Subcommittees at Macquarie University operate in accordance with the National Statement on Ethical Conduct in Human Research 2007, (updated July 2018), [Section 5.2.22].