Rise and fall of conflicting intuitions during reasoning

Bence Bago (bencebago@ed-alumni.net)
Paris Descartes University – Sorbonne Paris Cité
LaPsyDE (CNRS unit 8240)

Wim De Neys (wim.de-neys@parisdescartes.fr)
Paris Descartes University – Sorbonne Paris Cité
LaPsyDE (CNRS unit 8240)

Abstract

Recent dual process models proposed that the strength of competing intuitions determines reasoning performance. A key challenge at this point is to search for boundary conditions; identify cases in which the strength of different intuitions will be weaker/stronger. Therefore, we ran two studies with the two-response paradigm in which people are asked to give two answers to a given reasoning problem. We adopted base-rate problems in which base rate and stereotypic information can cue conflicting intuitions. By manipulating the information presentation order, we aimed to manipulate their saliency; and by that, indirectly the activation strength of the intuitions. Contrary to our expectation, we observed that the order manipulation had opposite effects in the initial and final response stages. We explain these results by taking into account that the strength of intuitions is not constant but changes over time; they have a peak, a growth, and a decay rate.

Keywords: reasoning; conflict detection; hybrid dual process model

Introduction

Decades of research in thinking and reasoning has revealed that people are usually subject to errors. Consider for example the following situation:

"There is a party with 1000 people. Jo is a randomly chosen participant from the party. We know that 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. We also know that 900 people attending the party are women. What is most likely: Is Jo a man or a woman?"

This is a so-called base rate problem. Based on the "normative"[1] principle that a randomly drawn individual will more likely come from the largest group, one should favor the conclusion that Jo is a woman. However, the majority of people tend to err on this problem by going with the presented stereotype (which cues that Jo is a man). Dual process theories provide an explanation for general thinking bias on problems such as the base rate task. They distinguish two types of processing, Type 1 and Type 2. One should note that there are many dual process theories, but in this study, we will focus on the most influential dual process theory, the default-interventionist theory. Type 1 processes (also referred to as intuitive processes) are thought to be completely autonomous, while Type 2 processes (also referred to as analytic processes) are more controlled. Type 1 processing generates responses cued by stereotypes or common beliefs; relying on this intuitive, initial response is what makes people biased in such situations. After Type 1 processing produced a response, in some cases, Type 2 processing gets engaged; this type of processing has the ability to override and correct the response generated by Type 1 processing. In general, it is assumed that Type 2 processing has the ability to generate responses based on logic or probabilities, while Type 1 processing has not been considered to be able to handle information such as logical properties of the task, or probabilities (Kahneman, 2011; Stanovich & Evans, 2013).

However, recently, conflict detection studies (De Neys, 2012, 2014) indicated that the assumption that Type 1 processing is not able to handle probabilistic or logical information might not hold. These studies showed that even biased reasoners were able to detect the conflict between intuitive "heuristic" cues (e.g., stereotypes) and "normative" logical and probabilistic principles (e.g., base rate probabilities). These studies usually contrast conflict and no-conflict reasoning problems. In conflict problems, heuristic processing and normative principles cue different responses as in the base rate problem above. In a no-conflict problem normative principles and heuristic processing cues the same response; for example, imagine that the above-presented base rate problem would state that there are 900 men and 100 women. In this case, both the stereotype and

[1] Note that we will be using the label "normative", ''correct", or ''logical" response as a handy shortcut to refer to ''the response that has traditionally been considered as correct or normative according to standard logic or probability theory". The appropriateness of these traditional norms has sometimes been questioned in the reasoning field (e.g.,see Stanovich & West, 2000, for a review). Under this interpretation, the heuristic response should not be labeled as ''incorrect" or ''biased". For the sake of simplicity, we stick to the traditional labeling. In the same vein, we use the term ''logical" as a general header to refer both to standard logic and probability theory.

base rate probabilities would cue the same response (that Jo is a man). In conflict problems, studies showed that even incorrect reasoners (compared to correct reasoners in no-conflict problems) showed elevated response times, decreased post-decision confidence, and higher activation in brain areas mediating conflict detection across a range of tasks (for review see De Neys, 2012).

These results made some authors suggest that there occurs some kind of elementary processing of logical/probabilistic information even during Type 1 processing. De Neys (2012) argues that conflict detection happens as a result of two conflicting Type 1 outputs, generated by two kinds of intuitions. He argues that one of these intuitions is based on stereotypes or common beliefs (heuristic intuition) the other one is based on logico-mathematical principles (logical intuition).

Recently, Bago and De Neys (2017a) went a step further and argued that people are not just able to detect the conflict intuitively but some of them are able to give the logically correct response intuitively. Our so-called hybrid dual process model argues that the two different intuitions differ in activation strength (or "salience"), and the actual intuitive response that the person provides will be the one which gained more strength. The relative difference between the strength of the heuristic and logical intuitions defines how pronounced the conflict is; the smaller the relative difference, the more pronounced the conflict will be; the larger the relative difference, the less pronounced it will be.

A key question at this point is to search for boundary conditions; identify cases in which the strength of different intuitions will be more or less pronounced. One way to do so is to manipulate the presentation order of base rate information and stereotypes. Let us explain why. In a previous study, Pennycook, Fugelsang, and Koehler (2015) argued that a "given piece of information is at its most salient just prior to judgement" (Pennycook et al., 2015, p. 57). Pennycook et al. (2015) further argued that this would mean that base rate information is most salient if presented right before the decision was made (after the stereotypical description had been presented). The authors observed that presenting the base rate information at the end of the problem indeed boosted participants' accuracy compared to the condition when it was presented first. To help us explain these results, one could operationalize saliency as the strength of a given intuitive response. Hence, whatever information was presented later, would be the more salient, therefore the intuition cued by this piece of information would be the stronger one.

In this study, we wanted to test the robustness of these findings – will we get the same effects after purely intuitive Type 1 processing? Thus, to test this question, one needs to use a research design which is able to separately measure intuitive Type 1 responses from analytic Type 2 responses. For this reason, we used the two response paradigm (Thompson, Prowse Turner, & Pennycook, 2011). In the two response paradigm, participants are presented with the same item twice. First, they are asked to give a very quick intuitive, initial response. Then, the same task is presented again and now they can take as much time as they want before providing their final response. One also needs to be sure that the initial response is truly intuitive; we achieved this by applying a strict response deadline (3 seconds) and a secondary task that burdens reasoner's (executive) cognitive capacity during the initial response. With these manipulations we can experimentally knock out Type 2 processing during the initial responding (Bago & De Neys, 2017a).

Our hypothesis was that if presentation order indeed affects the strength of an intuition, we should observe the same effect after purely intuitive processing as has been observed previously after deliberative thinking. That is, if base rates are presented last, the strength of the base rate intuition should be higher, and therefore more correct responses should be observed both at the initial and final response stages.

Study 1

Method

Participants

In total, 149 participants took part in the experiment (86 female, M = 39.3 year, SD =12.7 year). Participants were recruited online, via Crowdflower, and received $0.25 for their participation. Subjects were randomly assigned to one of the two conditions. Note that data in the S-BR condition were taken from the study of Bago & De Neys, (2017b). A total of 44.5% of participants reported having high school as highest completed educational level, while 52.1% reported that they have a post-secondary educational degree (3.4% reported less than high school).

Materials

Reasoning task. Participants solved a total of eight base-rate problems. All problems were taken from Pennycook, Cheyne, Barr, Koehler, and Fugelsang (2014). Participants always received a description of the composition of a sample (e.g., "This study contained I.T engineers and professional boxers"), base rate information (e.g., "There were 995 engineers and 5 professional boxers") and a description that was designed to cue a stereotypical association (e.g. "This person is strong"). Participants' task was to indicate to which group the person most likely belonged.

The problem presentation format we used in this research was based on Pennycook et al.'s (2014) rapid-response paradigm. In this paradigm, the base rates and descriptive information are presented serially and the amount of text that is presented on screen is minimized. Pennycook et al. introduced the paradigm to minimize the influence of reading times and get a purer and less noisy measure of reasoning time per se. Participants received 3 pieces of information in a given trial. First, the names of the two

groups in the sample (e.g., "This study contains clowns and accountants"). This sentence stayed on the screen and was always presented first. Participants were presented with stereotypical descriptive information (e.g., Person 'L' is funny) as well. The descriptive information specified a neutral name ('Person L') and a single word personality trait (e.g., "strong" or "funny") that was designed to trigger the stereotypical association. Participants also received the base rate probabilities. In this experiment, we manipulated the presentation order of the base rate probabilities and stereotypes. So, for one group the base rates were presented first (BR-S), for the other group, the base rates were presented last, after the stereotype (S-BR). Presentation order was manipulated between-subject. The following illustrates the full problem format in the S-BR condition:

This study contains clowns and accountants.
Person 'L' is funny.
There are 995 clowns and 5 accountants.
Is Person 'L' more likely to be:
o A clown
o An accountant

Half of the presented problems were conflict items and the other half were no-conflict items. In no-conflict items, the base rate probabilities and the stereotypic information cued the same response. In conflict items, the stereotypic information and the base rate probabilities cued different responses. Three kinds of base rates were used: 997/3, 996/4, 995/5.

Each problem started with the presentation of a fixation cross for 1000 ms. After the fixation cross disappeared, the sentence which specified the two groups appeared for 2000 ms. Then the first information appeared, for another 2000 ms, while the first sentence remained on the screen. Finally, the last information appeared together with the question and two response alternatives. Note that we presented the last information and question together (rather than presenting the last information for 2000 ms first) to minimize the possibility that some participants would start solving the problem during the presentation of the last part of the problem. Once all the parts were presented, participants were able to select their answer by clicking on it. The position of the correct answer alternative (i.e., first or second response option) was randomly determined for each item. The eight items were presented in random order.

Confidence in the correctness of the response was recorded after the initial and the final response stages by asking participants to indicate their confidence level on a scale ranging from 0% to 100%.

Cognitive load task. We used a concurrent load task - the dot memorization task - to burden participants' executive cognitive resources while they were solving the reasoning tasks. The idea behind the load manipulation is straightforward. One of the defining features of Type 2 processing is that it requires executive (working memory) resources (e.g., Evans & Stanovich,2013; Kahneman, 2011). Hence, if we burden participants' cognitive resources with a secondary load task while they are solving the reasoning problems, we reduce the possibility that they can engage in Type 2 thinking (De Neys, 2006).

In every trial, after the fixation cross disappeared, participants were shown a matrix in which 4 dots were presented in a complex interspersed pattern in a 3 x 3 grid for 2000 ms. Participants were instructed to memorize the pattern. Previous studies established that this demanding secondary task successfully burdens executive resources during reasoning (De Neys, 2006). After the matrix disappeared, the reasoning problem was presented as described above and participants had to give their first response. Then participants were shown four matrices with different dot patterns and they had to select the correct, to-be-memorized matrix. Participants were given feedback as to whether they recalled the correct matrix or not. Subsequently, the problem was presented again and participants selected their final response and response confidence. Hence, no load was imposed during the second, final response stage. All trials on which an incorrect matrix was selected (9.5 % of trials) were removed from the analysis.

Response deadline. In order to minimize the possibility of Type 2 engagement during the initial response, we used a strict response deadline (3000 milliseconds), based on a reading pre-test (see Bago & De Neys, 2017a). 1000 ms before the deadline, the background turned yellow to alert the participants to the approaching deadline. If participants did not select an answer within 3000 ms they got feedback to remind them that they had not answered within the deadline and they were told to make sure to respond faster on subsequent trials. Obviously, there was no response deadline on the final response, but only on the initial response. All trials where participants did not manage to provide a response were excluded from the analysis (8.7% of trials).

Procedure. The experiment was run online. People were clearly instructed that we were interested in their first, initial response to the problem. Instructions stressed that it was important to give the initial response as fast as possible and that participants could afterwards take additional time to reflect on their answer. After the instructions, participants were presented with practice problems to familiarize them with the procedure. At the end of the experiment, demographic questions were collected.

Results

Our main interest concerns the response accuracy analysis. Table 1 gives an overview of the findings. As one can see, we replicated the findings of Pennycook et al. (2015) at the final response stage for the conflict problems: Final accuracies on conflict problems are higher (41.6%) when the base rates are presented last vs. first (24.3%). However, contrary to our expectations, we do not observe the same effect at the initial response stage; there is even a trend towards fewer correct responses in the "base rates

last" S-BR condition (29.7%) vs BR-S (31.8%) condition. Indeed, the final conflict response accuracies in the S-BR condition were higher than the initial conflict response accuracies, whereas the reverse trend can be observed in the BR-S condition. In other words, the condition with the highest final accuracy (S-BR) was the one with the lowest initial accuracy, while the condition with the lowest final accuracy (BR-S) was the one with the highest initial accuracy.

Finally, as expected, note that accuracies on the no-conflict problems were always very high. Not surprisingly, in the absence of conflict, both the stereotype and base-rates can cue the correct response whatever order the information is presented in.

Table 1. Percentage of correct initial and final responses for conflict and no-conflict items in both order conditions.

	Response	Order	
		S-BR	BR-S
Conflict	Initial	29.7%	31.8%
	Final	41.6%	24.3%
No-conflict	Initial	93.4%	90.1%
	Final	93.7%	91.4%

Note. S-BR = base rates last/ BR-S = base rates first.

We used mixed effect logistic regression (logit) models to analyze the data and entered accuracy as a dependent variable. The order manipulation (S-BR/BR-S), response number (initial/final response), and their interaction were entered as predictors into the model. We also accounted for the random effect (random intercept) of subjects. We concentrated our analysis on the critical conflict problems. Only the interaction improved model fit significantly χ^2 (5) = 20.18, $p < 0.0001$, $b = 1.94$, but not the main effect of order χ^2 (3) = 0.19, $p = 0.66$ or response number χ^2 (4) = 0.38, $p = 0.54$. These results confirm our visual inspection that order affects initial and final accuracies differently.

Table 2. Frequency of each direction of change category (number of trials) for conflict items in both conditions.

Direction of change	Order	
	S-BR	BR-S
11	26.7% (54)	19.7% (47)
00	55.4% (112)	63.6% (152)
10	3% (6)	12.1% (29)
01	14.9% (30)	4.6% (11)

Note. S-BR = base rates last/ BR-S = base rates first.

For completeness, one could also test the direction of change in every trial (Bago & De Neys, 2017a). Specifically, people can give correct or incorrect responses on both response stages; this means that one could give two correct ("11"), two incorrect ("00"), an initial correct but final incorrect ("10"), or an initial incorrect but final correct ("01") response. The results of the direction of change analysis are summarized in Table 2. In both order conditions, the most frequent categories were the "00" and "11" cases. In line with previous observations (Bago & De Neys, 2017a; Thompson et al., 2011) people rarely changed their initial response (i.e., taken together the "10" and "01" cases account for 16%-18% of the trials). Interestingly, the direction in which people changed also tended to be reversed; in the S-BR condition most people who did change, changed from an incorrect to correct response (i.e., "01" category, 14.9% vs "10" category, 3%). However, in the BR-S condition most people who changed their initial response, changed it to an incorrect response (i.e., "10" category dominates with 12.1% vs 4.6% for the "01" category). Hence, this fits with the overall trend towards the higher likelihood of an initial incorrect and final correct response when the base rates are presented last. A Chi-square test of independence revealed that the distribution of the direction of change categories in the two order conditions significantly differed from each other χ^2 (3) = 27.56, $p < 0.0001$.

Discussion

Contrary to our expectations, we did not observe the expected accuracy effect at the initial response stage; we only observed it in the final response stage. However, we wanted to be sure that the findings were robust before drawing any conclusions. Note that Pennycook et al. (2015) already observed that their order findings were robust against manipulations of the extremity of the base rates. That is, they found the same order effect on (final) accuracies when they used so-called "moderate" base rates (e.g., base rate probabilities of 700 men and 300 women) instead of the "extreme" base rates (e.g., e.g. base rate probabilities of 995 men and 5 women) that were adopted in our (and their) Study 1. In Study 2 we therefore also adopted the moderate base-rates and examined whether the unexpected reversal of the order effect on initial, intuitive responses would still be observed.

Study 2

Method

Participants

In total, 162 participants took part in the experiment (98 female, M = 40.2 year, SD =14.6 year). Participants were recruited online, via Crowdflower, and received $0.25 for their participation. Subjects were randomly assigned to one of the two conditions. Note that data in the S-BR condition were taken from the study of Bago and De Neys (2017b). A total of 46.3% of participants reported having high school as highest completed educational level, while 52.5% reported that they have a post-secondary educational degree (1.3% reported less than high school).

Materials

Reasoning task. The identical experimental design was used as in Study 1. The only difference is that we used moderate base rates instead of extreme ones, namely 700/300, 710/290 and 720/280. In 16.7% of the trials participants did not provide the correct response for the dot matrix task, and in 10.5% of the trials, participants did not manage to produce an initial response within the deadline. These trials were excluded from further analysis. Overall, 24.6% of the trials were excluded and 977 were analyzed.

Results and discussion

Table 3 summarizes the accuracy results. As the table indicates, no-conflict response accuracies are again very high overall and we also replicated the conflict problem pattern we observed in Study 1: As Pennycook et al. (2015) found, presenting the base rates last led to increased accuracy on the final response. However, as in Study 1, the opposite trend was observed in the initial response. We also observe again that there were more initial than final incorrect response in the BR-S condition, whereas the opposite trend is observed in the S-BR condition. Statistical analysis on the conflict problems confirmed our visual inspection; neither presentation order χ^2 (3) = 0.04, p = 0.84, nor response number improved model fit significantly, χ^2 (4) = 0.05, p = 0.83, only their interaction did χ^2 (5) = 9.73, p = 0.0018, b = 1.4.

Table 3. Percentage of correct initial and final responses for conflict and no-conflict items in both order conditions.

	Response	Order	
		S-BR	BR-S
Conflict	Initial	16.4%	18.3%
	Final	23%	13.2%
No-conflict	Initial	90.9%	90.9%
	Final	90%	92.5%

Note. S-BR = base rates last/ BR-S = base rates first

Table 4 summarizes the results of the direction of change results for conflict items. Here too we observe the same trend as in Study 1. Among the few people who changed their response, the direction in which they changed are reversed as a function of presentation order; in the S-BR condition most people who did change, changed from an incorrect to correct response. But in the BR-S condition, more people changed to an incorrect response. A Chi-square test of independence revealed that the distribution of the direction of change categories in the two order conditions significantly differed from each other χ^2 (3) = 18.22, p = 0.0004.

Table 4. Frequency of each direction of change category (number of trials) for conflict items in both order conditions.

Direction of change	Order	
	S-BR	BR-S
11	14.2% (32)	8.2% (21)
00	74.8% (169)	76.7% (197)
10	2.2% (5)	10.1% (26)
01	8.8% (20)	5.5% (13)

Note. S-BR = base rates last/ BR-S = base rates first

General Discussion

In this paper, we tested whether manipulating the presentation order of the base rates and stereotypes had the same effect after purely intuitive processing (i.e., initial response) as had been observed previously after deliberative thinking (i.e., final response). In two studies, we replicated the findings of Pennycook et al. (2015) at the final response stage: Final accuracies on conflict problems were higher when the base rates were presented last. However, contrary to our expectations, in both studies this effect consistently reversed at the initial response stage. Why is this the case? We believe that these results draw attention to a simple but somewhat neglected issue in reasoning models, namely that intuitive responses are not generated instantly at full strength.

The hybrid dual process model that we presented in the introduction (e.g., Bago & De Neys, 2017a) argues that reasoning performance in the initial response stage is determined by the strength of different intuitions, for example. The implicit assumption here is that the strength of these intuitions is "instant" and "constant". That is, the idea is that the intuition is readily generated with full force and maintains this strength level.

However, upon some further reflection, this assumption might be quite naïve. It is reasonable to assume that even a quickly generated intuition needs some time to reach its peak. Keeping this feature in mind might suffice to explain the current findings. Have a look at Figure 1. In this illustration, the strength of two intuitions (I_1, I_2) change over time – they have a peak, a growth and a decay rate. The y-axis represents the strength, the x-axis represents time, while T1 and T2 represent the time of initial and final response, respectively.

I_1 and I_2 will start gaining strength when the relevant cue is presented (in the S-BR condition I_1 is the heuristic intuition cued by the presentation of the stereotype, and I_2 is the logical intuition cued by the base rate information). So, in the S-BR condition, the stereotype is presented first. When the stereotype is presented, the intuition (I_1) cued by it starts gaining strength. Subsequently, the presentation of the base rate information cues the logical intuition (I_2) and its strength will also start rising. Both intuitions grow until they reach their peak. At T1, I_1 has already reached its peak, and is stronger than I_2 (which has not reached its peak yet); as a result, I_1 will be the initial response. But after T1, the

strength of I_1 starts decaying, while the strength of I_2 is still increasing, and it reaches its peak at T2. At T2, I_2 will be the stronger intuition, so people will more likely pick I2 as their final response. Hence, the mere growth and decay of an intuition – or it's "rise and fall" as we labelled it in the title - implies that (ceteris paribus) the most recently cued intuition will be weaker earlier on in the reasoning process (e.g., initial response stage) and dominate later in the reasoning process (e.g., final response).

Clearly, we have presented and illustrated the most generic and general case in which two intuitions have the same peak level, growth, and decay rate. Obviously, these features might vary. One intuition might have a higher peak than the other, or a faster/slower growth/decay than the other. In addition, we believe that deliberation might also modulate the strength level. For example, one can imagine that one functional consequence of deliberation might be to boost or sustain the peak activation level of one intuition and decrease activation of the other. These more specific features have to be tested and validated in future studies. For example, one could try to test the role of deliberation by examining the impact of cognitive load on the presentation order findings in the second response stage. However, in all these more specific cases the general principle holds that we have to keep in mind that intuitions are not necessarily generated instantly but "rise and fall"; we need to consider their growth and decay. We believe this should motivate further research in the area by trying to determine what the growth and decay functions look like exactly.

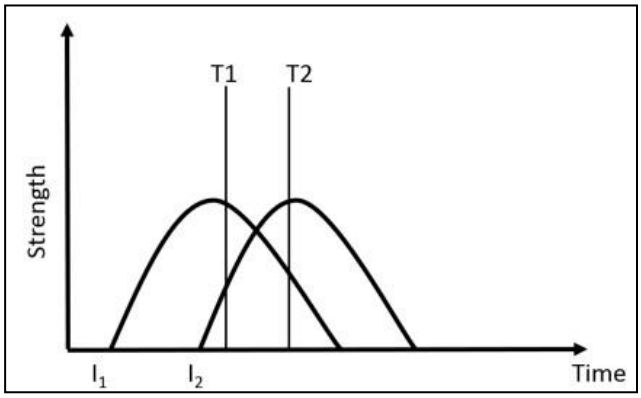

Figure 1. Illustration of how the strength of intuitions might change over time. The y-axis represents the activation strength while the x-axis represents time. I_1 and I_2 represent the two cued intuitions. Note that in the BR-S condition I_1 is the logical intuition cued by the base rate probabilities, while I_2 is the heuristic intuition cued by the stereotypes. Consequently, in the S-BR condition, I_1 is the heuristic and I_2 is the logical intuition. T_1 and T_2 represent the time of initial and final response, respectively.

Acknowledgement

Bence Bago was supported by a doctoral fellowship from Ecole des Neurosciences de Paris Ile-de-France. This research was also supported by a research grant (DIAGNOR, ANR-16-CE28-0010-01) from the Agence National de la Recherche.

References

Bago, B., & De Neys, W. (2017a). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109.

Bago, B., & De Neys, W. (2017b). Examining the hybrid dual process model. Manuscript in preparation.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, *59*(6), 1070–1100.

De Neys, W. (2012). Bias and conflict a case for logical intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38.

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

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)*, *54*(2), 128–133.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

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

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.

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.

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

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