Inhibition Failure is Mediated by a Disposition Toward Flexible Thinking

Alexander B. Swan (swan@psych.ucsb.edu)
Department of Psychological & Brain Sciences, University of California
Santa Barbara, CA 93106 USA

Russell Revlin (revlin@psych.ucsb.edu)
Department of Psychological & Brain Sciences, University of California
Santa Barbara, CA 93106 USA

Abstract

Conflict detection in dual process contexts is a widely studied phenomenon. However, only a small portion of the investigations has studied the role of individual differences in a typical conflict detection paradigm. In this study, participants completed a modified base-rate neglect task, as well as the Cognitive Reflection Task (CRT), and two Thinking Disposition Questionnaires. Results support an individual differences hypothesis in which the CRT prediction of accuracy on the base-rate conflict problems is mediated by the dispositional tendency to engage in flexible thinking.

Keywords: conflict detection, dual process, flexible thinking, individual differences

Introduction

A classic task in the biases and heuristics program begun by Tversky and Kahneman (1974) asks participants to make a judgment about group membership. A typical problem would look something like this (as appears in De Neys & Glumicic, 2008):

In a study 1000 people were tested. Among the participants there were 5 men and 995 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

In this example, the description of the individual is skewed heavily toward the stereotype associated with a man, even though in this sample, women drastically outnumber men. Due to the mismatch in cued responses (one response cued by the base-rates and another cued by the stereotypic description), the above problem represents a typical conflict problem. If the base-rates were flipped and the description remained the same, the same answer would be cued by both sources of information, and there would be no conflict (a nonconflict problem). This task is called the base-rate neglect task (Tversky & Kahneman, 1974). Overwhelmingly, people choose the stereotype answer when faced with a conflict problem, which is a result of applying the representativeness heuristic. In nonconflict problems, it doesn't matter what the reason for the choice is, since both the stereotype and base-rates are congruent (Tversky & Kahneman, 1974). Thus, error decisions made on this task are classified as a heuristic answer, since stereotypes fall within the "shortcut"/heuristic route when accessing stored information and making a decision. More recently, these findings and others have been explained by a group of theories called Dual Process Theories (DPT).

Dual process theories have a long and varied history in psychology (Frankish & Evans, 2009) with a common theme that human reasoning is characterized by a fast, automatic heuristic process (Type 1, or T1) and a slow, deliberate, analytic processing (Type 2, or T2) (e.g., De Neys, 2012; Evans & Stanovich, 2013; Frankish & Evans, 2009). And in the realm of normative logical reasoning, *heuristic* tends to be a pejorative term, usually indicating that a person has gotten the reasoning problem wrong, or anti-T2. Note, however, that a reasoner can also arrive at an incorrect answer using T2 processing (Evans, 2012; Stanovich, 2011; Thompson, Prowse Turner, & Pennycook, 2011). Due to the automaticity of T1 processing, it is argued to be the "default" state, and T2 is only engaged when it is needed (Evans, 2007).

De Neys and Glumicic (2008) have shown that the heuristics described by Tversky and Kahneman (1974) are consistently present and the data fit the DPT models well. Typically, the base-rates (or stereotypes) are chosen on nonconflict problems, but accuracy on conflict problems (choosing the base-rate) is dismally low. People prefer to choose the stereotype answer, and are usually slower on conflict problems when making this decision because it is regarded that the conflict within T1 (stereotype vs. base-rate) is detected and dealt with by engaging T2 to inhibit the stereotype answer (Pennycook & Thompson, 2012). In the latter situation, people are even slower when they get these conflict problems correct (Pennycook, Fugelsang, & Koehler, 2012).

Conflict Detection and Resolution in Dual Process Thinking

There are many models and theories proffered within the dual process approach and each describes the way a person might switch back and forth between T1 and T2 thinking. Clarification has been offered by the addition of two

concepts: *conflict detection*, in which the reasoner identifies that there is a conflict between heuristic and logical processing, *and resolution*, in which the reasoner decides which processes to apply (Evans, 2007).

The operation of the mechanisms of conflict detection and subsequent (or not) resolution is the open question in the DPT literature. Conflict arises when a solution to a given problem is found by T1, but T2 has found a different solution upon further reflection (De Neys, 2012, 2014). Represented this way, it appears that T2 only enters the equation when T1 has made an error. The reason for the error could be due to numerous apparent "shortcomings" of T1 processing, such as errant heuristics that do not apply to a given situation, processing that was too fast or incomplete to fully address the situation, overconfidence, or when an intuitive solution is patently false (De Neys & Glumicic, 2008; Evans & Curtis-Holmes, 2005; Thompson et al., 2011).

Recent evidence (De Neys & Franssens, 2009; De Neys & Glumicic, 2008; Pennycook & Thompson, 2012; Pennycook et al., 2012) points to the mechanism of conflict detection/resolution errors as a result of an inhibition failure. De Neys and Glumicic's (2008) findings were the first to support the *inhibition failure hypothesis*, wherein the stereotype answer is not inhibited on the conflict problems, and additional processing time is needed in order to fully inhibit the intuitive answer and achieve the "correct" answer (choose the base-rate answer).

Individual Differences in Conflict Detection and Resolution

Many of the previous studies discussed above (e.g., De Neys & Glumicic, 2008; Pennycook et al., 2012) investigated the conflict detection and resolution mechanism at the group level, describing what the average reasoner might do in a situation of T1/T2 conflict. While individual differences have not been fully neglected (e.g., Mevel et al., 2014; Pennycook et al., 2014), it is still an unresolved issue within DPT. While the mechanism of conflict detection and resolution might be universal, it is unclear which failures plague different types of reasoners. It may be fruitful to consider shifting to an individual-level analysis, where the behavior and processes of an individual are cast within the larger scope of DPT (De Neys, 2014).

In addition, individual differences are key to any good reasoning theory. Processes such as cognitive style, cognitive ability, and reasoning performance each potentially have something to add to the discussion of dual processing. It is possible that even the most proficient reasoners fail to choose the correct answer to a reasoning problem in a systematic way (Svedholm-Häkkinen, 2015)—why would that be? To answer this question a causal model of individual differences needs to be developed and tested. We note that performance on the Cognitive Reflection Test (CRT; Frederick, 2005), is positively correlated with performance on the conflict base-rate neglect problems. Both are direct behavioral measures of conflict detection and possible resolution. However, the CRT was designed to immediately measure inhibitory processes and behaviors. One of the classic problems is as follows: "A bat and a ball together cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?" The intuitive answer that is cued in the wording of the problem would be 10 cents, but this answer would be incorrect because of the "more than" phrasing. The correct answer, upon further reflection, is five cents. We contend that the inhibition failure hypothesis is the likely reason why people are biased on base-rate neglect problems (as opposed to conflict monitoring failures or storage failures). Thus, we argue that performance on the CRT is directly related to the propensity for choosing the base-rate. That is, the ability to inhibit the prepotent (intuitive) response on CRT word problems is the same ability that is needed to inhibit the salient stereotype information in the base-rate neglect problems.

We propose that it may take a specific type of thinking disposition (or cognitive style) to facilitate the inhibitory response described above. Although guessing on the CRT problems or mentally flipping a coin on the base-rate neglect problems can work to achieve the correct answer some of the time, performing well on these two tasks requires a desire on the part of the individual to engage T2 thinking and put forth the effort to correctly solve the problems. Two scales have been used widely in the DPT literature to assess these dispositions. The first is the Need for Cognition scale (NFC; Cacioppo, Petty, & Kao, 1984) that was designed to get a sense of how much an individual enjoys analytic thinking and engaging in difficult problem-solving. It would make sense that a person who desires engagement in thinking would use T2 processing to get a stronger score on the CRT and base-rate neglect tasks. The second commonly used thinking disposition scale in the DPT literature is the Actively Open-minded Thinking scale (AOT; Stanovich & West, 1997). This scale was created from a set of various related subscales. It was designed to measure a participant's sense of how open-minded they think they are (conversely, how cognitively rigid they are). More open-minded individuals tend to think about problems in a more flexible way, with an assumption that this is accompanied by T2 processing. These scales measure different aspects of the propensity to engage in effortful thinking, though they do correlate positively (Pennycook et al., 2014). A benefit of these two subjective measures is that they are independent of the inhibitory processes measured in the CRT and the base-rate neglect tasks.

Present Study

The present experiment was conducted to examine individual differences in conflict detection and resolution using established methodology. Specifically, a multiple mediation model was tested.

Base-rate neglect has been studied extensively in the last decade in the conflict detection realm of DPT. However, the methodology rarely has participants complete more than 20 base-rate problems. In this study, we expose participants to

50 trials of base-rate neglect problems in order to observe large-scale patterns. It is unknown whether the established methodology in DPT conflict detection and resolution can be replicated under conditions where participants are exposed to more than 20 base-rate problems. Since individual differences is the vehicle we are using to test the role of conflict detection in DPT, one way to test the reliability of a person's approach to the base-rate neglect problem is to expose them to a large amount of trials. In addition, repeated trials increase the sensitivity of the multiple mediation model.

The predictions for this experiment are as follows: (1) Performance of participants on conflict problems will be less accurate than performance on nonconflict problems, where the "correct answer" is choosing the base-rate on each problem, replicating recent base-rate neglect investigations (De Neys & Glumicic, 2008; Pennycook et al., 2012). (2a) Response times will be slower on conflict problems than on nonconflict problems, signaling the deliberation process of T2 engagement (De Neys & Glumicic, 2008). (2b) Response times on correctly answered conflict problems will be slower than on incorrectly answered conflict problems, indicating support for inhibition failure (De Neys & Bonnefon, 2013). These two behavioral predictions test the *inhibition failure hypothesis*.

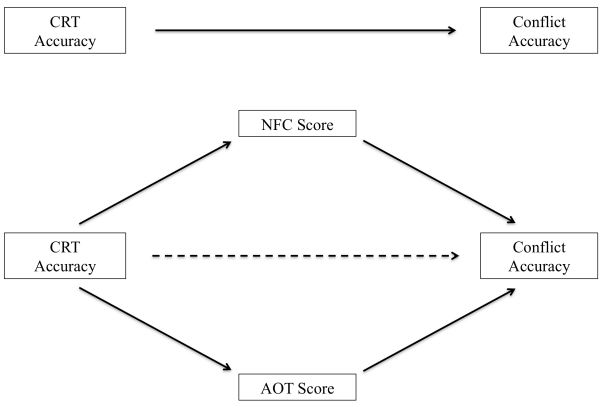

Figure 1: The proposed multiple mediation model. The dashed line in the second grouping indicates the reduced direct effect of CRT accuracy on conflict problem accuracy.

(3) The multiple mediation model tests an *individual differences hypothesis*: performance/accuracy on the CRT will positively predict performance on conflict problems. This is because both types of tasks require the participant to inhibit prepotent (intuitive) response. Moreover, this relationship is mediated by the disposition of the individual to engage in analytic and flexible thinking, which is correlated with engagement of T2 thinking. The two scales will be entered into a multiple mediation model as competing mediators to see if both, one, or neither scale mediates conflict problem performance. Put simply, CRT accuracy will predict accuracy on the conflict problems, and this relationship will be mediated by a higher disposition to engage in analytic and flexible thinking (see Figure 1 above).

Method

Seventy-six psychology undergraduate students (74% female), with a M_{age} = 18.54 years (SD = 1.01) participated in this study for partial course credit.

Each student was introduced to the study and told that there were four stages to the entire session. Participants first solved the three problems of the CRT. The CRT is a behavioral measure that tests a person's ability to inhibit an intuitive response on a word problem (Frederick, 2005). In addition to the bat-and-ball problem described earlier, there are two other word problems that have careful phrasing to elicit an intuitive answer that a person would have to inhibit in order to arrive at the correct answer. Using this behavioral method allows for an unbiased observation of cognitive reflectivity not achieved by some subjective measures.

Next, participants completed the base-rate neglect task. These were adapted from De Neys and Glumicic (2008). Participants answered 50 total base-rate problems, with 25 conflict problems and 25 nonconflict problems. The neutral condition from De Neys and Glumicic was not included to maximize the contrast in problems type (neutral problems do not offer stereotypic information).[1]

In addition to the congruency of the base-rate information and the stereotype information, there were three expressions of extreme base-rates (997 to 3, 996 to 4, and 995 to 5). This was done to vary the presentation of the problems, as well as to force reading of the base-rate information (it could be argued that if this information was the same that it would be ignored).[2] Previous research using this methodology has shown that the extreme base-rates are needed for a strong contrast between the conflict and nonconflict problems (De Neys & Glumicic, 2008; Pennycook et al., 2012). The order of these problems was fully randomized. Additionally, the answer was randomized, and it was either presented as the first option or second option (approximately 50% of the problems for each choice and for each problem type). Once the participant finished with those problems, they were given an opportunity to rest for 30 seconds.

Participants were then given the two thinking disposition questionnaires. These consisted of 18 items from the NFC scale (Cacioppo et al., 1984) and 41 items from the AOT scale (Stanovich & West, 1997). The NFC asked questions

[1] See De Neys and Glumicic (2008) for pretesting information regarding strength of stereotype for the problems tested. Stereotypes tested varied in content: age, gender, race, job-related groups, and stereotypical characteristics.

[2] De Neys and Glumicic (2008) performed pretesting on the extreme base-rate values to counter this argument and to vary the numbers in order to draw attention to differences between problems. Post-hoc analysis in their study showed that the small variation did not change performance.

to gauge the propensity of the participant to engage in effortful thinking (T2), such as "I prefer complex to simple problems." Participants rated their agreement with the statements on a five-point Likert scale, where larger values represented a characteristic quality of the individual and smaller values represented an uncharacteristic quality of the individual. The AOT was a composite questionnaire that gauged the cognitive flexibility of a person. In other words, it measures how willing an individual is to engage in effortful processing with information that has the potential to modify existing beliefs or evaluations. An example of a question from the AOT: "Difficulties can usually be overcome by thinking about the problem, rather than through waiting for good fortune." Participants rated their agreement on a six-point Likert scale, where larger values represented stronger agreement with the statement and smaller values represented stronger disagreement with the statement. Each scale had negatively-worded statements that were then reverse-coded (to prevent response acquiescence). Each question for each scale was randomized and the numerical scale appeared below each question.

Finally, a cognitive ability measure was gathered from participants. Used frequently in the cognitive literature (e.g., Stanovich & West, 2000), participants provided their most recent Scholastic Achievement Test (SAT) score (out of 2400). The majority of the undergraduates who participated in the study were from the western United States, so this is the likely standardized test taken prior to coming to college. Individual scores for subsections of the SAT were not sought.

Results

Table 1: Means for accuracy (proportion of base-rate answer) and response time (sec) for current experiment and for two related experiments with similar methodology.[3]

	Accuracy		Response Time	
Experiment	Con	Noncon	Con	Noncon
Present Study	0.48	0.90	12.5	11
De Neys & Glumicic (2008), Exps. 1 & 2	0.19	0.94	21	13
Pennycook et al. (2012), Exp. 3	0.59	0.95	14.1	11.5

Behavioral analyses were performed to corroborate previous research and replicate findings related to materials used (De Neys & Glumicic, 2008; Pennycook et al., 2012). Two main predictions were tested: (1) conflict problems would have lower accuracy scores (not choosing the base-rates) than nonconflict problems, and (2a and 2b) conflict problems would have slower response latencies than nonconflict problems, and correct conflict problems would have slower response latencies than incorrect conflict problems, reflecting the engagement of T2 thinking, or at least some deliberation (inhibition failure hypothesis).

Table 1 shows that Prediction 1 was supported: Participants chose the base-rate answer on conflict problems (M = .48, SD = .32) less often than on the nonconflict problems (M = .90, SD = .08), $t(75) = -13.56$, $p < .001$, Cohen's d = 1.40, which represents a departure from accurate responding (larger proportions here represent choosing the base-rate when it is the correct answer). Table 1 compares these findings to previous investigations of conflict detection in the base-rate neglect task, corroborating the presence of the conflict problem effect.

Again, Table 1 shows that Prediction 2a was supported. Participants took significantly longer to answer conflict problems (M = 12.5 s, SD = 4.27) than nonconflict problems (M = 11.0 s, SD = 3.47), $t(75) = 5.98$, $p < .001$, d = .36. However, Prediction 2b was not supported. When correct judgments of conflict problems were compared with incorrect judgments, correct judgments had slightly longer (M = 14.3 s, SD = 5.2) RTs than incorrect judgments (M = 13.9 s, SD = 4.8), but this difference was not reliable ($t(67)$ = .44, p = .33, d = .44, one-tailed; see Figure 2).[4] These results suggest that the conflict within these problems was detected, and deliberation did occur, but it is unclear whether inhibition failure was the culprit in errors on conflict problems, which is clearer in previous literature (e.g., De Neys & Glumicic, 2008).

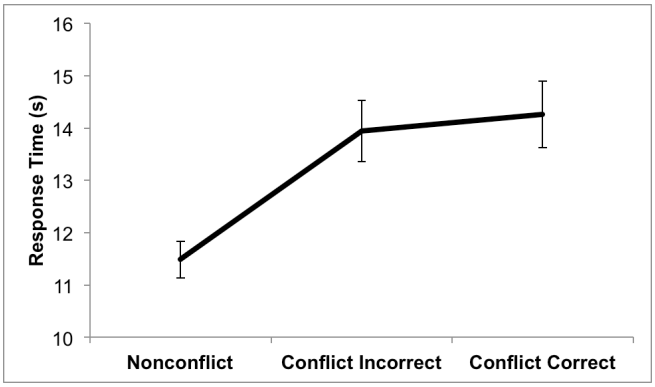

Figure 2: Response times for nonconflict problems and conflict problems, with conflict problems separated into whether the response was correct or incorrect.

Prior to testing the individual differences hypothesis (Prediction 3), inter-item reliability on the two thinking dispositions scales was performed. Both scales had good internal consistency: Cronbach's α = .84 for the NFC and .86 for the AOT. These values are similar to those found by Pennycook et al. (2014). Additionally, these two scales had a moderate positive correlation, $r(74) = .25$, $p = .03$.

The individual differences hypothesis stated that better

[3] RT values taken from Figure 3, De Neys & Glumicic (2008). They are approximate values.

[4] Eight participants were removed from this analysis because they had no incorrect conflict problem judgments.

performers on the CRT would opt to choose the base-rate more often on conflict problems. Also predicted was that this relationship would be mediated by the propensity to engage in analytic and flexible thinking, measured by the NFC and AOT scales. To test these hypotheses directly, we applied the Preacher and Hayes (2008) guidelines for bootstrapping in a multiple mediational regression analysis.

In the first equation, we regressed conflict problem accuracy on CRT accuracy. As predicted, better performers on the CRT tended to select the base-rate answer (the correct answer) on conflict problems ($\beta = .48$, $p < .001$). In the second equation, we regressed the thinking disposition scale scores as competing mediators on CRT accuracy. Both relationships were significant, whereby better performers on the CRT predicted the propensity to engage in general analytic thinking (NFC: $\beta = .36$, $p = .002$), as well as flexible thinking (AOT: $\beta = .32$, $p = .005$). In the third equation, conflict problem accuracy was regressed on CRT accuracy and the thinking disposition scales simultaneously. The model revealed that while flexible thinking, described by the AOT, was a significant predictor of the base-rate choice on conflict problems ($\beta = .27$, $p = .01$), a general propensity to engage in analytic thinking, measured by the NFC, was not a significant predictor on conflict problems ($\beta = .13$, $p = .24$). Moreover, the relationship between CRT accuracy and conflict problem accuracy was reduced from the first regression equation ($\beta = .35$, $p = .002$). A Sobel test was conducted to determine if the amount of mediation was significant; the indirect effect was significant ($z = 2.29$, $p = .02$). Bootstrapping (at 2000 resamples with replacement; Preacher & Hayes, 2008) confirmed the Sobel test with a 95% CI [.034, .131]. Since the direct path from CRT accuracy to conflict problem accuracy remained significant, we conclude that that this relationship is only partially mediated by flexible thinking. Figure 3 illustrates the mediation model with beta weights for each path.

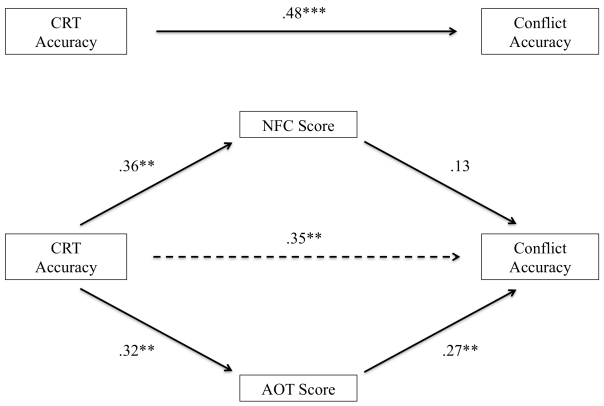

Figure 3: Mediation model displaying the causal paths for each of the variables. Paths are notated by their beta weights. ***$p < .001$, **$p < .01$.

Taken together, these findings are consistent with the hypothesis that the ability to inhibit the prepotent heuristic response on the CRT is similar to inhibitory processing in the base-rate neglect paradigm, and that this relationship in processing occurs because of the disposition of the individual to be a flexible thinker when presented with a judgment problem, but not a product of a general propensity to think analytically. Furthermore, these effects also suggest support for the inhibition failure hypothesis in conflict detection, wherein an individual who is a more flexible thinker can suppress the intuitive response on both the CRT and the base-rate neglect task.

Discussion

The goal of the present study was to replicate and extend the current literature of conflict detection and resolution using the base-rate neglect paradigm and to add a new causal model to the individual differences aspect in DPT investigations (see Figure 3).

In a small modification of the existing methodology of the base-rate neglect task, similar behavioral performance was observed (Table 1), increasing the validity of the overall findings. In addition, some support for the inhibition failure account through response latencies was observed. Though the time difference between correct conflict problems and incorrect conflict problems was not reliable, the medium effect size is promising for future studies. Perhaps response times decreased as a function of trials, since participants were exposed to many problems. Additional data analysis will be performed to investigate this possibility. The results show that processing times increase when there is an informational conflict within the problem. Furthermore, even more time is needed to properly deliberate and engage T2 thinking in order to choose the correct answer. As Table 1 suggests, incorrect responses are the result of a failure to inhibit the salience of the stereotype answer, and the longer response latency associated with this is indicative of the overall conflict detection (cf. De Neys & Glumicic, 2008).

A multiple mediation model was tested, adding a causal account to the discussion of individual differences and support for the inhibition failure hypothesis. Accuracy on the CRT predicts subsequent accuracy on conflict base-rate problems and this relationship is partially mediated by a flexible thinking disposition. Similarly, Pennycook et al. (2014) argued that the AOT is a better indicator of willingness to engage in analytic thinking (flexibility) and that the NFC is merely a broadly-defined tendency to think analytically in the realm of conflict detection. The results of our mediational analysis support the idea that the AOT and NFC measure two separate aspects of analytic thinking, as the AOT predicted performance on conflict problems whereas the NFC did not (cf. Svedholm and Lindeman, 2013).

While this individual differences model shows a meaningful relationship between accuracy on the CRT, accuracy on the conflict problems, and a flexible thinking disposition, it is important to interpret these conclusions

with caution. Individual difference measures are difficult to separate from other intrinsic motivations a person might have, but the use of two related but separate scales was an attempt to overcome this limitation and be more comprehensive. However, other factors should not be ignored. For example, cognitive ability has been shown to be a strong measure of performance in dual process literature (Stanovich & West, 2000). Stanovich (2011) argues that T2 thinking is the center for general intelligence and cognitive ability, and people with higher cognitive ability have an easier time operating in T2. SAT scores were gathered for this study, but were not included in the main analyses or mediation model because initial correlations indicated that SAT scores were only marginally related to accuracy on conflict problems ($r(72) = .22$, $p < .10$). Additional measures of cognitive ability are needed to form a better picture of its relationship to conflict detection and resolution.

Acknowledgments

The authors would like to thank Alexander Hill for data collection for this project and four anonymous reviewers for their helpful comments and suggestions.

References

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of Need for Cognition. *Journal of Personality Assessment*, *48*(3), 306–307.

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*, 169–187.

De Neys, W., & Bonnefon, J.-F. (2013). The "whys" and "whens" of individual differences in thinking biases. *Trends in Cognitive Sciences*, *17*(4), 172–178.

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

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

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

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, *18*(1), 5–31.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389.

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

Frankish, K., & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual process and beyond* (pp. 1–29). Oxford: Oxford University Press.

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

Mevel, K., Poirel, N., Rossi, S., Cassotti, M., Simon, G., Houdé, O., & De Neys, W. (2014). Bias detection: Response confidence evidence for conflict sensitivity in the ratio bias task. *Journal of Cognitive Psychology*.

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. (2012). Are we good at detecting conflict during reasoning? *Cognition*, *124*(1), 101–106.

Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic Bulletin & Review*, *19*(3), 528–534.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891.

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

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342–357.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–65; discussion 665–726.

Svedholm, A. M., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology, 104*(3), 303-319.

Svedholm-Häkkinen, A. M. (2015). Highly reflective reasoners show no signs of belief inhibition. *Acta Psychologica*, *154*, 69–76.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–40.

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