2019-10-02

The logic sense: exploring the role of executive functioning in belief and logic-based judgments

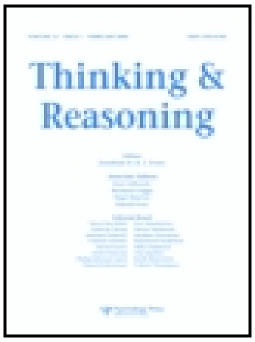
Howarth, S

http://hdl.handle.net/10026.1/12954

10.1080/13546783.2018.1523808
Thinking and Reasoning
Taylor & Francis (Routledge)

Thinking & Reasoning

ISSN: 1354-6783 (Print) 1464-0708 (Online) Journal homepage: http://www.tandfonline.com/loi/ptar20



The logic sense: exploring the role of executive functioning in belief and logic-based judgments

Stephanie Howarth, Simon Handley & Clare Walsh

To cite this article: Stephanie Howarth, Simon Handley & Clare Walsh (2018): The logic sense: exploring the role of executive functioning in belief and logic-based judgments, Thinking & Reasoning, DOI: 10.1080/13546783.2018.1523808

To link to this article: https://doi.org/10.1080/13546783.2018.1523808

Published online: 25 Oct 2018.

THINKING & REASONING
https://doi.org/10.1080/13546783.2018.1523808

The logic sense: exploring the role of executive functioning in belief and logic-based judgments

Stephanie Howarth[a], Simon Handley[a] and Clare Walsh[b]

[a]Department of Cognitive Science, Macquarie University, Sydney, Australia; [b]School of Psychology, Plymouth University, Plymouth, UK

ABSTRACT

The Default Interventionist account suggests that by default, we often generate belief-based responses when reasoning and find it difficult to draw the logical inference. Recent research, however, shows that in some instances belief judgments take longer, are more prone to error and are more affected by cognitive load. One interpretation is that some logical inferences are available automatically and require intervention in order to respond according to beliefs. In two experiments, we investigate the effortful nature of belief judgments and the automaticity of logical inferences by increasing the inhibitory demands of the task. Participants were instructed to judge conclusion validity, believability and either font colour or font style, to increase the number of competing responses. Results showed that conflict more strongly affects judgments of believability than validity and when inhibitory demands are increased, the validity of an argument impacts more on belief judgments. These findings align with the new Parallel Processing model of belief bias.

ARTICLE HISTORY Received 26 November 2017; Accepted 9 September 2018

KEYWORDS Belief bias; intuitive logic; conditional reasoning; dual process theory; inhibition; parallel processing

A dominant view of human reasoning is that our thinking is beset by error and bias and that logical thought can only be accomplished through deliberative and effortful processing. According to this view, our beliefs often have a profound impact on our capacity to reason logically, providing powerful, pre-emptive cues to a solution that comes to mind rapidly and requires effort to resist in favour of more considered responding. The psychological study of belief bias in reasoning has contributed much to this model of thinking and a dual-process account that has dominated both the academic (Evans, 2007, 2008, 2010b) and popular scientific literatures during the last decade (Evans, 2010a; Kahneman, 2011).

Typical studies of belief bias involve presenting participants with reasoning problems in which there is a conflict between the believability of the

CONTACT Stephanie Howarth stephanie.howarth@mq.edu.au

conclusion and its logical status. Consider, for example, the following problem:

All humans are organisms

All organisms are made of plastic

Therefore, all humans are made of plastic.

Participants are usually asked to assume the premises are true and to evaluate whether the conclusion logically follows. In this case, the argument is valid, but the conclusion is unbelievable. Generally, people find this type of "conflict" problem more difficult than logically equivalent problems where there is no such conflict, presumably because a belief- based response must be resisted prior to a logical judgment being realised. This type of explanation for the differential difficulty of problem types is captured by Default Interventionist (DI) Dual Process Accounts of thinking (see, for example, Evans & Stanovich, 2013). Such accounts define two types of process; fast, implicit, default type-1 processes which often deliver responses based on beliefs or learned associations; and type-2 processes which are characteristically slower, explicit by nature and draw upon working memory (WM) or executive processes. Controlled processing is required to inhibit responses based upon type-1 processes and provide the cognitive resources to reason logically. However, a number of authors have argued that humans have an inclination towards "miserly" processing (Toplak, West & Stanovich, 2011) and given the cognitive effort required to inhibit a belief-based response (Handley, Capon, Beveridge, Dennis & Evans, 2004), we are often inclined to default to type-1 processing due to its low computational expense (Toplak, West & Stanovich, 2014).

There is some support for the claim that belief-based errors reflect superficial, default processing whereas logical responding depends upon more effortful reflection. De Neys (2006), for example, used a secondary task to load working memory while participants completed syllogistic reasoning problems in which logic and belief-based responses diverged. He showed that a working memory load impacted accuracy on conflict items, but not on items where there was no-conflict between logic and belief. The findings suggest, in line with DI accounts, that cognitive resources are required to resist default responses and reason logically. Similarly, when participants are given limited time to respond, belief-based responses are increased on syllogistic reasoning tasks (Evans & Curtis-Holmes, 2005) and more invalid inferences are drawn on conditional inference tasks (Schroyens, Schaeken & Handley, 2003). Studies of individual differences have shown that logical accuracy on tasks in which there are competing heuristic and normative responses is associated with variations in cognitive ability and rational thinking dispositions (Stanovich, 1999). These correlations are assumed to arise because they index variations in the capacity to engage in inhibitory processes which are required to resist the lure of a compelling and highly available heuristic response (Stanovich & West, 2008). This has been identified as the central feature of type-2 processing necessary for hypothetical thinking (Evans, 2007).

Default Interventionist accounts assume that belief-based responses are available early and the logical response requires further deliberative processing. Recent research, however, offers evidence of fast (and slow) processing for both rule-based and belief-based reasoning (Newman, Gibb & Thompson, 2017). Furthermore, reasoners are aware of early conflict even when a biased response is generated. De Neys (2012) has argued that individuals automatically detect logic-belief conflict, even when the conflict is resolved in favour of belief-based responding. Conflict problems lead to increased processing latencies compared to non-conflict problems (Stupple & Ball, 2008) and are associated with lower levels of reported confidence (De Neys, Cromheeke & Osman, 2011). Bago and De Neys (2017) further supported this claim by testing the time course assumption of the Default Interventionist account, using a two-response paradigm requiring immediate and delayed responding to base rate and syllogistic reasoning problems. They found that reasoners showed a high prevalence of logical responses to conflict problems even under immediate responding conditions and accompanying load, consistent with an intuitive route to logical responding. Participants were also less confident in the accuracy of their answers under these conditions suggesting the simultaneous and intuitive generation of both logical and heuristic (belief-based) responding.

The detection of conflict is also reflected in neurophysiological measurement, whereby the processing of conflict problems is associated with increased skin conductance and activation of the Anterior Cingulate Cortex (De Neys, Moyens & Vansteenwegen, 2010). Interestingly, the detection of conflict is not influenced by secondary tasks or cognitive capacity as one might expect if the logical processing involved depended upon explicit deliberative reasoning (De Neys et al., 2010, 2011; De Neys & Glumicic, 2008).

Recent research suggests that intuitive sensitivity to logical structure arises because logical arguments are more fluent; that is, it is easier to integrate the premises of a logically valid argument than one that is not (Morsanyi & Handley, 2012). Increased fluency is thought to lead to increased positive affect which has been shown to result in higher ratings of liking for valid than invalid arguments (Morsanyi & Handley, 2012; Trippas, Handley, Verde & Morsanyi, 2016). Most recently, it has been shown that the conclusions of valid arguments are judged to be brighter on a perceptual judgment task, presumably because a feeling of positive affect is misattributed to the perceptual features of a valid conclusion that follows fluently from its premises (Trippas et al. 2016).

How might we reconcile the apparent evidence for effortful processing in the resolution of belief-logic conflict and the mounting evidence for intuitive detection of logical structure? A key feature of studies that provide support for the DI account is that in all cases participants are instructed explicitly to reason logically. That is, they are instructed to consider what logically follows from a set of premises independent of their belief state. One possibility is that responding to instructions where there is a conflict between the instructional requirement and beliefs draws upon cognitive resources, so although a logical response may be available intuitively, any available response must be validated against the task requirements. The nature of the instructions provided has been the focus of a number of recent studies in which participants were instructed to either judge the logical validity of a presented conclusion or its believability (Handley, Newstead & Trippas, 2011). Handley, Newstead, and Trippas (2011) adapted the typical belief-bias paradigm by asking participants to judge both the validity and the believability of an argument. Contrary to predictions derived from the DI account, belief judgments took longer and were more subject to error than judgments of logical validity, and belief-logic conflict had a greater impact on the accuracy of belief than logic judgments. They concluded that the logical inference was available early and had to be inhibited in order to explicitly evaluate a conclusion's believability. One possible interpretation of these findings is that the logical inference is available automatically and by default and intervention is required to successfully give a belief judgment. Howarth, Handley, and Walsh (2016) tested this account by examining the impact of a secondary task (Random Number Generation) on both belief- and logic-based judgments while completing both simple (Modus Ponens) and more complex (Disjunctions) reasoning problems. The results replicated earlier findings showing that belief-based judgments produced lower rates of accuracy overall and were influenced to a greater extent than validity judgments by the presence of a conflict between belief and logic for both simple and complex arguments. However, the secondary task, while reducing accuracy of judgments overall, had its greatest impact on logic-based judgments. Howarth et al. (2016) interpreted their findings as a conflict between two type-2 processes. They propose that logical responses are available at an early type-1 level, yet require type-2 processing in order to explicitly extract the underlying structure of an argument required for delivering a response based on the validity of the inference. Belief-based responses, on the other hand, are available later and effective responding depends upon the *inhibition* of an accessible

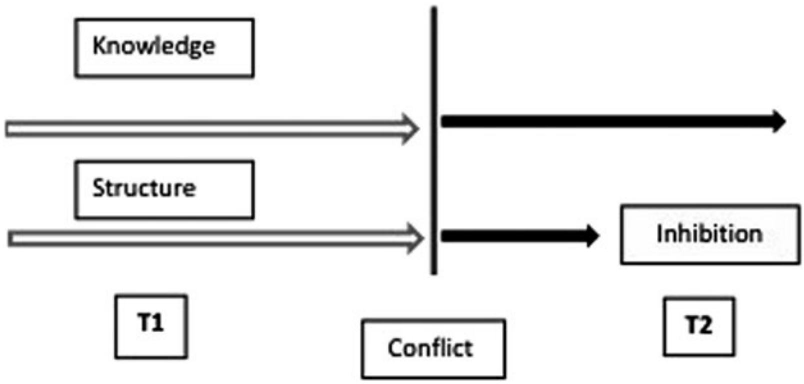


Figure 1. The parallel competitive (PC) dual-process model where processing time is represented on the right to left dimension, the vertical line indicates the point where conflict might be detected and also the transition between type-1 and type-2 processing. For simple logical problems, a response based upon logical structure is available first and must be inhibited in order to respond on the basis of beliefs.

and competing intuitive logical response. This interpretation is consistent with the parallel processing model introduced by Handley and Trippas (2015) as illustrated in Figure 1.

This model suggests that both structural features and knowledge of problem content can be triggered concurrently. Both can demand type-1 and type-2 processing; likewise, both are prone to errors and biased responses. Notably, one of the key characteristics of this model is that the direction of interference between knowledge based and structural processing will depend upon the complexity of the logical argument, such that simple problems may cue type 1, logical responses early which will require inhibiting in order to produce a knowledge-based output.

Howarth et al. (2016) argued that both belief- and logic-based judgments depend upon type-1 and type-2 processes, where interference can occur bi-directionally at the type-2 stage or type-1 stage of processing which accounts for the evidence of intuitive detection of conflict and the impact of conflict under explicit instruction. This model has subsequently received support from evidence which shows that the direction of the impact of conflict on logic and belief judgments depends upon the complexity of the logical task (Trippas, Thompson & Handley, 2017). With more complex logical arguments, such as multiple model syllogisms, conflict leads to reduced logical performance and has a limited effect on belief judgments, but the reverse holds with simpler logical structures, such as conditional or disjunctive arguments. These findings align with the substantial evidence for belief bias in the literature and for the idea that this varies as a function of inhibitory capacity (Handley et al., 2004; De Neys & Franssens, 2009).

The primary aim of this article is to investigate the effortful nature of belief-based judgments in line with the model detailed in Figure 1. If a simple logical inference is available early, then the ability to respond on the basis of belief would require the inhibition of the readily available logical response. One way to investigate this account is to increase the inhibitory demands of the task. This was achieved by increasing the number of available competing responses on the task through the introduction of a third instructional condition. Our parallel competitive model predicts that for simple logical problems, an inhibitory load will increase the difficulty of withholding a logical response and will therefore reduce accuracy on the belief judgments.

Experiment 1

In Experiment 1, we presented conflict and no-conflict problems under three competing instructional conditions. Participants responded based upon instructional cues that required judgments of conclusion validity, conclusion believability, or the font colour of the final word of the conclusion. Consider, for example, the following problem:

If the fruit is a strawberry, then it is purple

Suppose the fruit is a strawberry

Does it follow that the fruit is red?

Participants were then instructed to respond in line with one of three response criteria:

1. Logic: Valid/Invalid
2. Belief: Believable/Unbelievable
3. Colour: Red/purple

In the Stroop condition, on belief/logic conflict trials, there was a mismatch between the meaning of the final word and its colour, where the automatic inclination to respond based on the written word is inconsistent with the colour of the ink (Stroop, 1935). This mismatch was designed to increase the inhibitory demands of the task and create a Stroop effect in the Stroop condition, a feature that was not present in the control condition, for example:

Either blood is red or white

Suppose the blood is not white

Does it follow that blood is red?

We predicted that the introduction of an additional demanding instructional condition would have its greatest impact on belief judgments under conflict.

Method

Participants

A total of 112 psychology undergraduate students from Plymouth University took part in Experiment 1, in exchange for two course credits. Ninety-two women and 20 men participated in the experiment and were randomly assigned to either the Control or the Stroop condition.

Design

A 3(Instruction) × 2(Problem Type) × 2(Complexity) × 2(Condition) mixed design with repeated measures on the first three factors, required participants to judge the validity or believability of the conclusion or the font colour of the last word in the conclusion, creating three *instructional* sets. Participants were presented with both conflict and no-conflict *problem types*, which refers to the conflict between a belief and a logical response. The *complexity* of the argument refers to the simple Modus Ponens form or the more complex Disjunctive form and *condition* refers to the Stroop or Control group as the between subject factor.

Materials

A set of 192 logical arguments were created for Experiment 1. As in previous work (Howarth et al. 2016), there were an equal number of Modus Ponens and Disjunctive arguments. Half the problems were conflict problems (belief and logic conflicted) and half were no-conflict problems (belief and logic matched).

The third instructional cue was principally introduced to create the Stroop condition which would increase the inhibitory demands of the reasoning task and allow us to measure the impact this would have on belief and logic judgments. To achieve this, the third instruction was an adaptation of the Stroop task, which involved colour naming of a concluding word. In the Stroop condition, under colour naming instruction, the trials were a mix of congruent and incongruent trials. This condition was designed to reflect the original design of the Stroop task where there are a mix of congruent and incongruent trials. This increases inhibitory demands because individuals are unaware from trial to trial whether the colour word and ink colour will be in conflict with one another or not. For incongruent

Table 1. Examples of response options for each instructional cue with conflict and no-conflict items in the Experimental condition.

Instructional cue	No-conflict problems	Conflict problems
	If the fruit is a strawberry, then it is red Suppose fruit is a strawberry Does it follow that the fruit is red?	If the fruit is a strawberry, then it is purple Suppose fruit is a strawberry Does it follow that the fruit is red?
	Either flamingos are pink or they are purple Suppose flamingos are not purple Does it follow that flamingos are pink?	Either flamingos are pink or they are purple Suppose flamingos are not pink Does it follow that flamingos are not purple?
Colour naming – MP	Red ☑ or Purple ☒	Purple ☑ or Red ☒
Colour naming - Disjunctive	Pink ☑ or Purple☒	Pink ☑ or Purple ☒
Belief	Believable ☑ or Unbelievable ☒	Believable ☑ or Unbelievable ☒
Logic	Valid ☑ or Invalid ☒	Valid ☒ or Invalid ☑

trials, there was a mismatch between the ink colour and the written word. For congruent trials, the ink colour and word matched (Table 1). In the control condition, participants were also presented with the colour naming task; however, all the trials were congruent.

In both the Stroop and control conditions, participants were also presented with belief and logic problems; and, from trial to trial, would not know how to respond until presented with response options. When required to make a belief judgement, participants were presented with "believable" or "unbelievable" as response options, when required to answer according to logic, they were presented with "valid" or "invalid" response options.

In the control condition, the correct response always matched the font colour and the written word on every trial.

Since we were interested in the impact of conflict on belief and logic judgments, we ensured that the incongruent colour naming trials (colour mismatch) were also conflict problems (conflict between belief and logic) and hypothesised that the embedded colour incongruence would increase inhibitory demands for conflict problems under belief and logic instruction. In the control condition, there was a match between the word and colour ensuring no increased inhibitory demands on conflict problems.

Procedure

Participants were tested in groups of up to four (maximum) individuals in partitioned booths and were randomly assigned to the Stroop or Control condition. In both conditions, each participant was presented with 192 problems with an optional respite at the half way point. The experiment was administered through a computer-based E-prime programme. Participants were first required to answer 12 practice trials before being

Table 2. Average accuracy and latency scores on conflict and no-conflict items, across complexity, under colour-naming instruction for both conditions.

Variable	Conflict		No-conflict		Overall means
	MP	Disj	MP	Disj	
Stroop					
Response Accuracy (%)	89(.01)	98(.01)	99(.01)	100(.01)	97(.01)
Latency (ms)	4191(162)	3971(173)	4201(169)	4099(160)	4116(143)
Control					
Response Accuracy (%)	99(.01)	97(.01)	98(.01)	97(.01)	98(.01)
Latency (ms)	2893(169)	2801(180)	3323(176)	3146(167)	3041(149)
Mean Accuracy (%) (across each cell)	94(.01)	98(.01)	99(.00)	99(.01)	
Mean Latency (ms) (across each cell)	3542(117)	3386(125)	3762(122)	3623(115)	

Results exclude below chance scores and include correct only latencies. Stroop condition ($N = 50$), Control condition ($N = 46$).
Note: standard error in brackets.

presented with the experimental trials in an unblocked design. The randomisation of the trials ensured that participants would remain unaware as to whether they would need to answer according to logic, beliefs or the font colour of the last word, until the response options were presented. For each problem, the first premise was presented alone on the computer screen for a total of 3,000ms. It remained on the screen and was followed by the second premise, conclusion, and response options, which were all presented together. The full problem stayed on the screen until the participant pressed their response key on the keyboard. Both accuracy scores and latencies were recorded for each trial.

Results

Prior to analysing the data, we adopted the same procedure as Handley, Newstead, & Trippas (2011; see also Howarth et al., 2016) and eliminated participants that scored below 50% on conflict items under both belief and logic instruction as this indicates a failure to engage with the instructions. A total of 16 participants were eliminated, three from the Stroop condition ($N = 50$) and 13 from the Control condition ($N = 46$).

From Tables 2 and 3, accuracy on colour naming trials and no-conflict belief and logic trials was near ceiling. Ceiling effects can have a significant impact on homogeneity of variance resulting in higher rates of skewness and kurtosis across conditions. The normality of the data was initially checked on non-transformed data, across both conflict and non-conflict

Table 3. Average accuracy and latency scores for belief and logic instruction on conflict and no-conflict problems, for MP and disjunctive arguments.

Variable	Belief conflict		Belief no-conflict		Logic conflict		Logic no-conflict		Overall means
	MP	Disj	MP	Disj	MP	Disj	MP	Disj	
Stroop									
Response Accuracy (%)	82 (.03)	82 (.03)	97 (.01)	90 (.02)	95 (.01)	90 (.02)	97 (.01)	91 (.02)	91 (.01)
Latency (ms)	6018 (278)	6126 (275)	5375 (221)	5917 (287)	6473 (268)	6628 (330)	5426 (239)	5778 (233)	5968 (225)
Control									
Response Accuracy (%)	88 (.03)	85 (.03)	97 (.01)	90 (.02)	92 (.02)	85 (.02)	98 (.01)	91 (.02)	91 (.01)
Latency (ms)	5754 (290)	6235 (287)	5155 (230)	5876 (299)	5560 (279)	6326 (344)	5245 (249)	6023 (243)	5772 (236)
Mean Accuracy (%) (across each cell)	85 (.02)	83 (.02)	97 (.01)	90 (.01)	94 (.01)	88 (.02)	98 (.01)	91 (.01)	
Mean Latency (ms) (across each cell)	5886 (201)	6181 (199)	5265 (160)	5897 (207)	6016 (194)	6477 (238)	5335 (173)	5901 (166)	

Results exclude below chance scores and include correct only latencies. Stroop condition ($N = 50$), Control condition ($N = 46$).
Note: standard error in brackets.

problems for belief and logic instructions, by examining measures of skewness and kurtosis for each cell of the design.

Prior to transformation, the z scores across conditions ranged between −6.0 and −8.9 for skewness and between −2.6 and 10.3 for kurtosis, as one might expect given the observed ceiling effects in certain cells. To control for the ceiling effects associated with these high accuracy rates, particularly on no-conflict problems, an arcsine transformation was carried out to improve homogeneity of variance (Milligan, 1987). The arcsine transformation brought the z-scores closer to normal range [within ±2.58, see Hair, Black, Babin, Anderson & Tatham, 2006; Mat Roni, 2014) with skewness reduced to a range between −0.5 and −3.7 and kurtosis ranging between −0.01 and 2.00. Although the transformation did not completely eliminate skewness, it created a more homogenous data set on which to perform the analysis of variance. This choice of transformation was in keeping with previous work (Handley et al. 2011; Howarth et al. 2016). On the latency data, the analysis was conducted on correct responses where outliers more than two standard deviations from the mean of each cell were eliminated and any missing data were replaced with the overall cell mean, which accounted for no more than 5.5% of the overall data. Latencies and accuracy scores are presented in the tables prior to transformation.

We separate out the analysis and start with the colour naming instruction to determine (1) whether we had created the desired Stroop effect in the Stroop condition and that this condition was sufficiently more demanding than the control condition and (2) whether a colour mismatch (incongruent trials) resulted in less accuracy than colour match (congruent) trials. Then, we carry out separate analyses on belief and logic instruction in order to determine whether the difference between inhibitory demanding and less demanding conditions differentially impacted on belief and logic judgments and evaluated problem conflict through accuracy and latency data.

Colour- naming instruction

A problem type (Conflict/No-conflict) by complexity (MP/Disjunctives) by experimental condition (Stroop/Control) mixed design ANOVA was carried out on Arcsine transformed accuracy data under colour-naming instruction (Table 2). Problem type indicates whether a conflict or no-conflict problem accompanies a colour-naming trial. Recall that the control condition only included congruent trials whilst the Stroop condition contained a mix of congruent and incongruent trials. The incongruent trials always accompanied conflict problems.

The results revealed a main effect of *problem* type (Conflict/No-conflict); $F(1, 94) = 37.066$, $p < .001$, $\eta_p^2 = .283$, with colour naming trials accompanying conflict problems producing lower accuracy scores than those

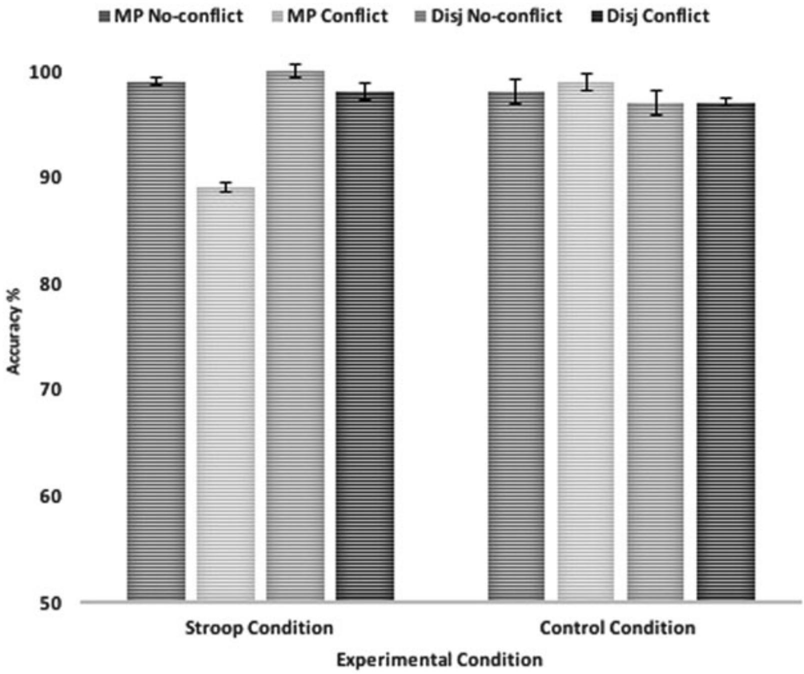


Figure 2. An illustration of the differential impact of conflict on MP and Disjunctives arguments under each condition. Error bars represent the SEM.

accompanying no-conflict problems *(96% vs. 99%)* and a main effect of *condition;* $F(2, 94) = 5.845$, $p = .018$, $\eta_p^2 = .059$, which highlighted poorer performance in the Stroop condition, substantiated by a significant interaction between *problem* type and *condition*; $F(1, 94) = 54.942$, $p < .001$, $\eta_p^2 = .369$. Separate analyses revealed that the main effect of *problem* type was only present in the Stroop condition*;* $F(1, 49) = 93.371$, $p < .001$, $\eta_p^2 = .656$, showing poorer performance on colour-naming trials accompanying conflict problems compared to no-conflict problems *(94% vs. 100%)*. This demonstrates that the presence of a colour mismatch on conflict problems in the Stroop condition leads to reduced accuracy, confirming the presence of increased inhibitory demands and the desired Stroop effect.

There was also a significant main effect of *complexity* (Disjunctives/MP); $F(1, 94) = 34.873$, $p < .001$, $\eta_p^2 = .271$, showing lower accuracy scores on colour-naming trials accompanying MP arguments compared to disjunctives arguments *(96% vs. 99%).* Although the difference is small, the effect suggests that trials where the inference is accomplished rapidly (MP) interferes with the capacity to correctly name the colour of the last word. This significant effect of complexity also interacted with condition; $F(1, 94) = 115.982$, $p < .001, \eta_p^2 = .552$. Furthermore, the 3-way interaction between *problem* type, *complexity* and *condition*; $F(1, 94) = 69.692$, $p < .001$, $\eta_p^2 = .426$, was driven by the significantly larger interaction between *problem* and *complexity* for the Stroop condition; $F(1, 49) = 106.408$, $p < .001$, $\eta_p^2 = .685$, compared to the control condition; $F(1, 45) = 5.447$, $p = .024$, $\eta_p^2 = .108$ (see Figure 2).

The results revealed that conflict had a bigger effect on colour-naming trials accompanying MP arguments *(99% no-conflict vs. 89% conflict)* compared to those accompanying disjunctives arguments *(100% no-conflict vs. 98% conflict)* in the Stroop condition. In the control condition, there was no effect of conflict on disjunctive judgments *(97% no-conflict vs. 97% conflict) or MP* judgments *(98% no-conflict vs. 99% conflict),* under colour-naming instruction. Suggesting that when you have incongruent inhibitory demanding trials, colour naming is more difficult, particularly when it is a readily available inference (MP).

The latency data for colour-naming instruction produced a main effect of *problem* type; $F(1, 94) = 7.145$, $p = .009$, $\eta_p^2 = .071$, with colour-naming trials accompanying no-conflict problems taking longer than conflict problems *(3,692ms vs. 3,464ms).* This could be due to the fact that no-conflict problems strongly cue a response that is compatible with both belief and logic, and therefore, it takes longer to shift to a response based on colour. There was a main effect of *condition*; $F(2, 94) = 27.137$, $p < .001$, $\eta_p^2 = .224$, indicating that overall, performance in the Stroop condition was significantly slower than the control condition *(4,116ms vs. 3,041ms),* consistent with the Stroop condition being more demanding on inhibition. Finally, there was a small main effect of *complexity;* $F(1, 94) = 5.400$, $p = .022$, $\eta_p^2 = .054$, showing that under colour-naming instruction arguments presented as MP took slightly longer than those presented as disjunctives *(3,652ms vs. 3,504ms).*

Belief and logic instruction

Separating out the analysis on colour-naming items allowed us to show that the Stroop manipulation was effective in increasing performance demands in the experimental (Stroop) condition. Next, we examine the experimental manipulation on belief and logic instruction, given that the form and content of the arguments are critical features in determining responses under these conditions. We report analysis on both subjects and items in order to ensure that all the effects are robust across the range of contents used and that subject effects are not linked to a particular set of items in the reasoning task.

The presence of a subject's effect in the absence of an items effect suggests that the finding is not generalisable across the full item set. An items effect in the absence of a subject's effect suggest that the finding is not generalisable across the full subject sample. The observation of an effect

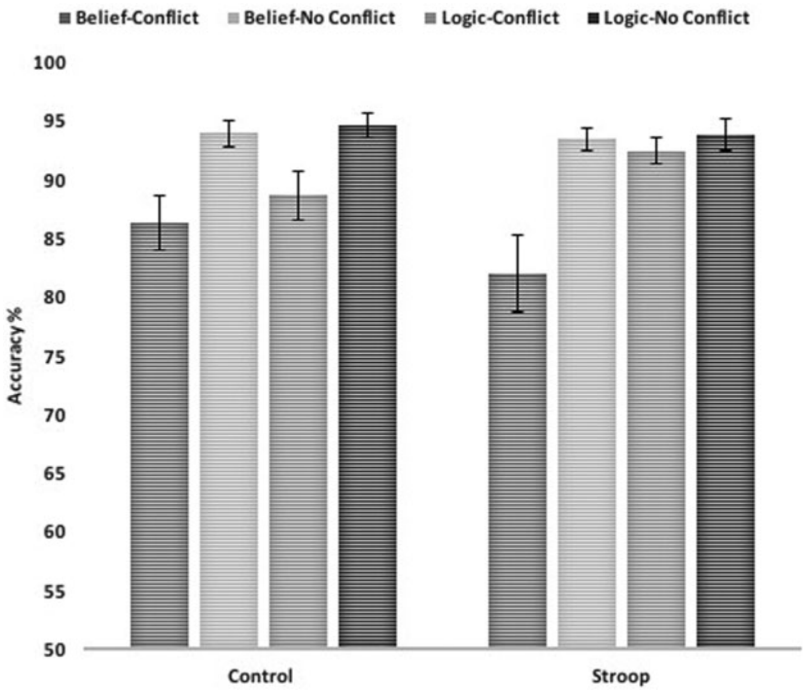


Figure 3. An illustration of the differential impact conflict has on belief and logic judgments in each condition. Error bars represent the SEM.

that holds across subjects and items indicates a robust finding that is not dependent upon specific item contents or on specific participant subsets. We consequently only follow up effects of this kind that allow some generalisability.

The analysis is presented as $F_{1(subjects)}$ and $F_{2(items)}$ and the accuracy data were analysed using a 2(Belief/Logic) × 2(Conflict/No-conflict) × 2(MP/Disjunctives) × 2(Stroop condition/Control condition) mixed design ANOVA on Arcsine transformed data (Table 3). The analysis showed a main effect of *instruction* (Belief/Logic) for subjects, F_1 (1, 94) = 22.637, $p < .001$, $\eta_p^2 = .194$; and items, F_2(1, 120) = 11.749, $p = .001$, $\eta_p^2 = .089$, with logic judgments generating higher accuracy scores compared to belief judgments *(93% vs. 89%)*. There was a main effect of *problem* type (Conflict/No-Conflict) for subjects, F_1 (1, 94**)** = 56.043, $p < .001$, $\eta_p^2 = .374$; and items, F_2(1, 120) = 37.594, $p < .001$, $\eta_p^2 = .239$, with poorer performance on conflict items compared to no-conflict items *(88% vs. 94%)*. There was a main effect of *complexity* (Disjunctives/MP) for subjects, F_1 1, 94) = 50.773, $p < .001$, $\eta_p^2 = .351$; and items, F_2(1, 120) = 31.355, $p < .001$, $\eta_p^2 = .207$, where MP produced higher accuracy scores than disjunctive judgments *(94% vs. 88%),* but there was no main effect of *condition* for subjects, F_1(2, 94) = .010, $p = .919$, $\eta_p^2 < .001$; or items, $F_1(1, 120) = .257$, $p = .613$, $\eta_p^2 = .002$, showing that performance between the Stroop and control condition was not significantly different *(91% vs. 91%)*.

There was a significant interaction between *instruction* and *problem* type for subjects, $F_1(1, 94) = 8.467$, $p = .004$, $\eta_p^2 = .083$; and items, $F_2(1, 120) = 6.994$, $p = .009$, $\eta_p^2 = .055$, revealing that the belief/logic conflict had more of an impact on judgments of the conclusions believability *(85% B-conflict vs. 94% B-no-conflict)* than its logical validity *(91% L-conflict vs. 95% L-no-conflict)*. There was also a marginal 3-way interaction between *instruction, problem* type and *condition* for subjects, $F_1(2, 94) = 3.710$, $p = .057$, $\eta_p^2 = .038$; and a significant interaction for items, $F_2(1, 120) = 20.516$, $p < .001$, $\eta_p^2 = .146$. (Figure 3).

A follow-up analysis on both conditions revealed that the interaction between *instruction* and *problem* was present in the Stroop condition for subjects, $F_1(1, 49) = 8.871$, $p = .004$, $\eta_p^2 = .153$; and items, $F_2(1, 120) = 19.384$, $p < .001$, $\eta_p^2 = .139$, but absent in the control condition for both subjects, $F_1(1, 45) = .789$, $p = .379$, $\eta_p^2 = .017$; and items, $F_2(1, 120) = .223$, $p = .637$, $\eta_p^2 = .002$. This demonstrated that the effect of conflict was enhanced for belief instruction *(82% conflict vs. 94% no-conflict)* compared to logic instruction *(92% conflict vs. 94% no-conflict)* in the Stroop condition. In the control condition, the effect of conflict was comparable between belief instruction (*86% conflict vs. 94% no-conflict)* and logic instruction *(89% conflict vs. 96% no-conflict)*. A significant interaction between *problem* type and *complexity* for subjects, $F_1(1, 94) = 17.322$, $p < .001$, $\eta_p^2 = .156$, suggested that the impact of conflict was larger for MP judgments *(89% conflict vs. 97% no-conflict)* than disjunctives *(86% conflict vs. 91% no-conflict)* (Table 3). However, this interaction was not significant in the items analysis; $F_2(1, 120) = 2.271$, $p = .134$, $\eta_p^2 = .019$. Finally, there was a 3-way interaction between *instruction, problem* and *complexity* for subjects, $F_1(1, 94) = 6.131$ $p = .015$, $\eta_p^2 = .061$; however, the interaction was not significant for the items analysis, $F_2(1, 120) = 1.070$, $p = .303$, $\eta_p^2 = .009$.

A mixed design ANOVA on response latencies, produced no main effect of *instruction* for both subjects, $F_1(1, 94) = 2.150$, $p = .146$, $\eta_p^2 = .022$; and items, $F_2(1, 120) = .027$, $p = .870$, $\eta_p^2 < .001$, revealing no significant difference in latencies between belief *(5,807ms)* and logic *(5,933ms)* instruction. There was, however, a main effect of *problem* type for subjects, $F_1(1, 94) = 34.661$ $p < .001$, $\eta_p^2 = .26$; and items, $F_2(1, 120) = 5.879$, $p = .017$, $\eta_p^2 = .047$, with conflict items taking longer to complete than no-conflict items (*6,140ms vs. 5,600ms*) and a main effect of *complexity* for subjects, $F_1(1, 94) = 38.171$, $p < .001$, $\eta_p^2 = .289$; and items, $F_2(1, 120) = 8.895$, $p = .003$, $\eta_p^2 = .069$, indicating faster response latencies to MP judgments compared to disjunctive judgments (*5,626ms vs. 6,114ms*). The interaction between *complexity* and *condition* for subjects, $F_1(1, 94) = 6.329$, $p = .014$, $\eta_p^2 = .063$; and items, $F_2(1, 120) = 5.048$, $p = .026$, $\eta_p^2 = .040$, suggested that there was bigger effect of condition on MP latencies (*5,428ms control vs. 5,823ms Stroop)*, compared to the latencies on disjunctive judgments (*6,115ms control vs. 6,112ms Stroop)*. The effect of experimental *condition* was not significant in the subject's analysis, $F_1(2, 94) = .359$, $p = .550$, $\eta_p^2 = .004$; but was for the item's analysis, $F_2(1, 120) = 16.339$, $p < .001$, $\eta_p^2 = .120$, with items in the Stroop condition taking longer to complete *(5,772ms control vs. 5,968ms Stroop)*.

Lastly, there was a significant 3-way interaction between *instruction*, *problem* type and *condition* in the subject's analysis, $F_1(1, 94) = 7.290$, $p = .008$, $\eta_p^2 = .072$; which was not significant for the item's analysis, $F_2(1, 120) = 2.404$, $p = .124$, $\eta_p^2 = .020$.

Discussion

In Experiment 1, we asked participants to evaluate the validity or believability of a given conclusion on simple (MP) and complex (disjunctives) items, whilst introducing a third instructional colour-naming condition designed to increase the inhibitory demands of the task and allow us to evaluate the role of inhibition in the effect of conflict on belief-based judgments. We conjectured that an increase in inhibitory demands would have its greatest impact on belief judgments; in line with this prediction the findings demonstrated that the effect of conflict increased under belief instruction in the Stroop condition. In other words, logic judgments appeared to impact on belief judgments more when inhibitory demands were increased by the inclusion of an arduous, third instructional cue.

Importantly, the experiment also clearly replicated earlier findings by Howarth et al. (2016) demonstrating that belief judgments are more prone to errors, conflict problems produce lower accuracy scores and MP judgments produce high accuracy scores. Furthermore, results reproduced the now familiar finding where belief-logic conflict impacts more on the believability of the conclusion for both accuracy and latency scores (Handley et al., 2011; Howarth et al., 2016). In addition, the items analysis showed that the effects on instruction can be extended beyond the present sample and beyond the present set of items.

The current findings lend support to the parallel processing model outlined in Figure 1. According to the model, simple logical inferences are available early and require inhibiting in order to produce a belief-based output. Limiting the capacity to inhibit the early completion of a logical response, by the inclusion of an additional instructional cue, should have its greatest impact on belief judgments, as confirmed in Experiment 1.

However, while the findings suggest that belief judgments are more effortful when the number of competing responses are increased; the Stroop design appears to have a more impact on the length of time taken to complete logic-conflict items (*6501ms*) compared to belief-conflict items (*6072ms*). We will consider the reason why this is the case in more detail in the general discussion but in short, we conjecture that this can be explained in terms of there being "Two Routes" to a logical output. One route is an independent type-1 process completing first and creating an intuitive cue, possibly accompanied by a feeling of rightness (Thompson, Prowse Turner & Pennycook, 2011), based on the logical structure of the argument. This output is what requires inhibiting and causes a type-1/type-2 conflict when instructed to reason on the basis of beliefs. The second route to a logical solution is a type-2 process, activated when given explicit instructions to reason logically, that runs parallel to the route required for a belief-based response. In Experiment 1, the second route may be impacted by the increased inhibitory demands created by the third instructional cue which interferes with the length of time it takes for a logical output to complete. However, these logical inferences are still simple and are therefore less prone to error than belief-based judgments.

At this point, we consider two possible explanations for the increased inhibitory demands associated with the Stroop condition: (1) due to the nature of the Stroop instruction and the requirement to inhibit semantic processing in favour of colour-naming or (2) because the presence of a third instructional condition with significant processing demands increases the number of competing and effortful responses, thus enhancing the inhibitory demands of the task. In order to discriminate between these two explanations, in Experiment 2, we introduced a third instructional condition in which the cue directly taxes on WM (n-back) as opposed to inhibition. If the observed findings are the result of competing responses created by an additional instructional cue, then we can expect to see an impact of the n-back task on belief judgments.

Experiment 2

The aim of the second experiment was to determine whether the nature of the processing demands of the third instructional condition was responsible for the findings in Experiment 1. In Experiment 2, we examined the function of memory updating in relation to logic and belief-based judgments. Memory updating is the executive function that monitors incoming information for its magnitude and appropriateness, and updates the information in WM by exchanging old items for new (Morris & Jones, 1990). In the current experiment, the colour-naming instructional condition is replaced with *n*-back instructions which involve the identification of features presented in a previous trial (see design and materials sections). The *n*-back (Kirchner, 1958) is a performance task requiring continuous maintaining, updating, and releasing of arbitrary bindings between items and temporal order positions (Friedman et al. 2006). The third instructional cue in the current experiment is designed so participants are unable to predict whether they will be required to make a judgment based on believability, validity or recall the features in the preceding trial. Unlike the colour-naming task, the *n*-back manipulation should not draw directly on inhibitory resources. Therefore, if the findings in Experiment 1 are a result of the specific inhibitory demands of the colour-naming instruction then any impact on belief judgments should be eliminated. However, if the effects arise because of the presence of an additional effortful competing response option, then we would expect to replicate the findings of Experiment 1.

Method

Participants

A total of 79 Psychology undergraduate students from Plymouth University took part in the current study, in exchange for course credits or £8 payment. Fifty-nine females and 20 male participants were randomly assigned to either the Experimental (memory updating) or Control condition (matching).

Design

As in Experiment 1, a 3(instruction) × 2(Problem Type) × 2(Complexity) × 2(Condition) mixed design was used with repeated measures on the first three factors. Participants were instructed to make judgments on the validity or believability of a conclusion or on the font style of the last word in the previous trial, creating the three *instructional* sets. Again, participants were presented with both conflict and no-conflict problems (problem type) in the form of MP and disjunctive arguments (complexity) in either the Experimental or Control group (condition) as the between subject factor.

Materials

The same 192 arguments used in Experiment 1 were used for Experiment 2 with the 64 colour-naming trials exchanged for 64 *n*-back trials in the experimental condition and 64 matching trials in the control condition. The allocation of items to participants was carried out in the same way as Experiment 1, where half of the 192 items were conflict problems (belief

Table 4. Examples of response options for each instructional cue in both the Experimental (n-back) and Control (matching) conditions.

Instructional cue	Trial one	Trial two
	Either flamingos are pink or they are purple Suppose flamingos are pink Does it follow that flamingos are Purple?	If the bird is a Dove then it is orange Suppose the bird is a Dove Does it follow that the bird is Orange?
n-back (experimental)	n/a (no previous trial)	Same ☒ or Different ☑
Belief	Believable ☒ or Unbelievable ☑	Believable ☒ or Unbelievable ☑
Logic	Valid ☑ or Invalid ☒	Valid ☑ or Invalid ☒
Matching (control)	Same ☒ or Different ☑	Same ☑ or Different ☒

and logic conflicted) and half were no-conflict problems (belief and logic matched). Of the conflict items, half had valid/unbelievable conclusions and the other half were invalid/believable. Of the no-conflict items, half had valid/believable conclusions, and the other half had invalid/unbelievable conclusions to ensure the equal distribution of yes/no responses under each instructional condition. Finally, half of the items were disjunctive arguments and half Modus Ponens arguments.

In the experimental condition, the *n*-back trials were created by changing the font style of the last word in the conclusion and participants were required to indicate whether the font of the last word in the previous trial matched the font of the current trial by responding "same" or "different". In the control condition, participants received 64 matching trials in place of the *n*-back trials and were required to match the font style of the last word in premise one with the font style of the last word in the conclusion also by responding "same" or "different". Participants in both the Experimental and Control condition received half "same" and half "different" items. The following five font styles were used in size 18 font: Lucida Handwriting, Bradley Hand, Algerian, *Ravie* and Curlz MT.

Again, in both the experimental and control condition, participants were also presented with belief and logic problems and from trial to trial, would not know how to respond until presented with response options (see Table 4).

Procedure

Participants were tested in maximum groups of 4, randomly allocated to the Control or Experimental condition. Each participant was presented with 192 problems with an optional respite at the half way point: 64 were presented under belief instruction (believable/unbelievable) 64 under logic instruction (Valid/Invalid) 64 under *n*-back instruction in the experimental condition (same/different) and 64 under matching instruction in the control condition (same/different). There were 12 practice trials to begin followed by the full 192 trials which were presented in one of 16 unique orders. The

Table 5. Average accuracy and latency scores on conflict and no-conflict problems, across complexity, under n-back instruction (experimental) and matching instruction (control).

Variable	Conflict		No-conflict		Overall means
	MP	Disj	MP	Disj	
Experimental Condition					
Response Accuracy (%)	81(.03)	76(.03)	81(.03)	84(.02)	80(.02)
Latency (ms)	3265(227)	3411(211)	3007(203)	3324(257)	3252 (197)
Control condition					
Response accuracy (%)	92(.03)	85(.03)	93(.02)	88(.03)	90(.02)
Latency (ms)	3973(230)	4504(214)	3960(205)	4721(260)	4289 (200)
Mean Accuracy (%) (across each cell)	86(.02)	80(.02)	87(.02)	86(.02)	
Mean Latency (ms) (across each cell)	3619(161)	3958(151)	3484(144)	4023(183)	

Results exclude below chance scores and include correct only latencies: Experimental condition ($N = 37$), Control condition ($N = 36$).
Note: standard error in brackets.

n-back instructions emphasised the importance of remembering the characteristics of the previous trial, while the control task restricted the matching of font style to the current trial. Trials where presented for the same amount of time and in the same format as Experiment 1 and both accuracy scores and latencies were recorded for each trial.

Results

Two participants were eliminated from the Experimental condition ($N = 37$) and four from the Control condition ($N = 36$), for performing below chance on conflict items. Consistent with Experiment 1, normality was checked on non-transformed accuracy data, by examining measures of skewness and kurtosis. The data were then Arcsine transformed to control for celling effects and improve homogeneity of variance. Prior to transformation, the z scores across conditions ranged between −6.9 and −8.1 for skewness and between 6.1 and 10.9 for kurtosis. Post–transformation, the data were more normally distributed with z-scores ranging between −2.3 and −3.4 for skewness and 0.2 and 1.18 for kurtosis. Analysis on the latency data were conducted on correct responses where outliers were eliminated and missing data accounted for no more than 5.0% of the overall data.

Once again, we present the analyses for the *n*-back and matching instruction first, in order to examine the demands of the experimental

Table 6. Average accuracy and latency scores for belief and logic Instruction on conflict and no-conflict problems, for MP and disjunctive arguments.

Variable	Belief Conflict		Belief No-conflict		Logic Conflict		Logic No-conflict		Overall Means
	MP	Disj	MP	Disj	MP	Disj	MP	Disj	
Experimental									
Response Accuracy (%)	80 (.04)	84 (.03)	98 (.01)	91 (.02)	96 (0.2)	91 (.03)	99 (.02)	89 (.03)	91 (.02)
Latency (ms)	6615 (349)	6318 (367)	6244 (288)	5971 (264)	6464 (258)	6880 (278)	6290 (226)	6832 (282)	6452 (234)
Control									
Response Accuracy (%)	84 (.04)	86 (.03)	93 (.01)	90 (.02)	92 (.02)	85 (.03)	95 (.02)	89 (.03)	89 (.02)
Latency (ms)	6446 (353)	6336 (372)	5717 (292)	6090 (268)	5993 (261)	6371 (281)	5358 (229)	6366 (286)	6085 (237)
Mean Accuracy (%) (across each cell)	82 (.03)	85 (.02)	96 (.01)	91 (.02)	94 (.02)	88 (.02)	97 (.01)	89 (.02)	
Mean Latency (ms) (across each cell)	6531 (248)	6327 (261)	5981 (205)	6031 (188)	6229 (183)	6626 (198)	5824 (161)	6599 (201)	

Results exclude below chance scores and include correct only latencies. Experimental condition ($N = 37$), Control condition ($N = 36$).

manipulation compared to the control condition, followed by separate analyses on belief and logic instruction.

N-back vs. matching instruction

A 2(Conflict/No-conflict) × 2(MP/Disjunctives) × 2(Experimental/Control Condition) mixed design ANOVA was carried out on Arcsine transformed accuracy data under *n*-back instruction (Table 5). As in Experiment 1, *problem* type refers to whether a conflict or no-conflict problem accompanies the *n*-back and matching trials. The results uncovered a marginal main effect of *problem* type (Conflict/No-conflict); $F(1, 71) = 3.827$, $p = .054$, $\eta_p^2 = 0.051$, showing that the *n*-back and matching trials accompanying conflict items produced marginally lower scores than no-conflict items *(83% vs. 87%)*. There was main effect of *condition;* $F(2, 71) = 17.132$, $p < .001$, $\eta_p^2 = 0.194$, which confirmed that performance was lower in the experimental condition than the control condition *(80% vs. 90%)*. This shows that the requirement to remember the font characteristic in the previous trial significantly increased the processing demands and difficulty of the task. There was also a main effect of *complexity* (Disjunctives/MP); $F(1, 71) = 12.712$, $p = .001$, $\eta_p^2 = 0.152$, where *n*-back and matching trials accompanying disjunctives produced lower accuracy scores than those accompanying MP *(83% vs. 87%)*, suggesting that participants found it more challenging to recall or match font style when the problem was more complex (Table 5). Contrasting with the Stroop task (Experiment 1), where rapidly accomplished MP inferences interfere with the ability to inhibit the colour name, *n*-back trials appear to tax distinct executive functions, supposedly those more demanding of working memory, thus having more impact on inferences requiring additional cognitive effort to resolve (there were no significant interactions to report; all p's $> .05$).

Response latencies for *n*-back instruction produced no main effect of *problem* type; $F(1, 71) = 0.144$, $p = .705$, $\eta_p^2 = 0.002$, but there was a main effect of *condition*; $F(2, 71) = 13.691$, $p < .001$, $\eta_p^2 = 0.162$, showing that performance in the control condition was significantly slower than in the experimental condition *(4289ms vs. 3252ms)*. There was also a main effect of *complexity;* $F(1, 71) = 22.106$, $p < .001$, $\eta_p^2 = 0.237$, showing that *n*-back and matching trials accompanying disjunctive arguments took longer to complete than MP arguments *(3,991ms-Disjunctives vs. 3552ms-MP)*. The interaction between *complexity* and *condition;* $F(1, 71) = 4.934$, $p = 0.030$, $\eta_p^2 = 0.065$, indicated that *n*-back performance on disjunctive judgments took longer than MP judgments in the control condition (*4613ms-Disjunctives vs. 3967ms-MP*) compared to the experimental condition (*3368ms-Disjunctives vs. 3136ms-MP*), but the effect size was small. All remaining interactions were not significant (all p's $> .1$).

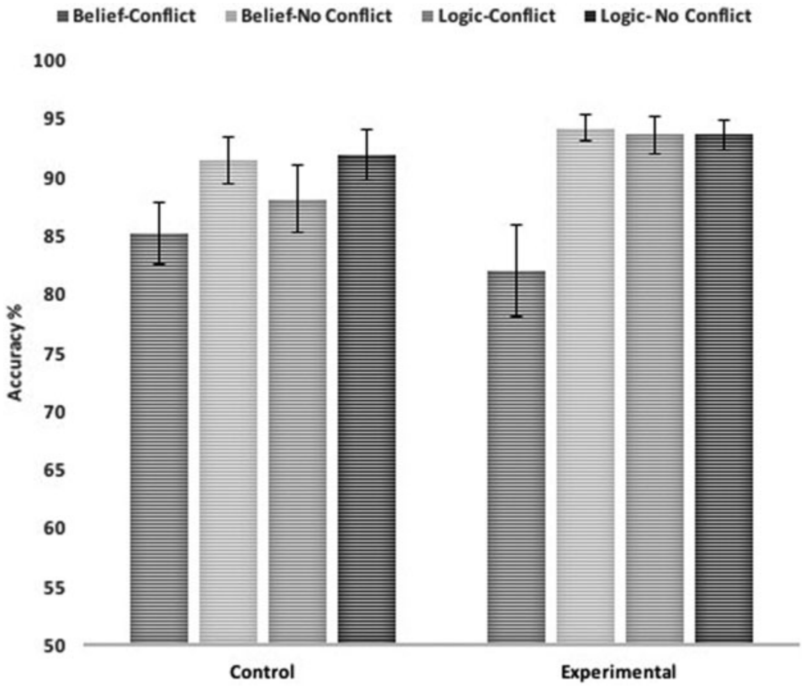


Figure 4. An illustration of the differential impact conflict has on belief and logic judgments in each condition. Error bars represent the SEM.

Belief and logic instruction

A 2(Belief/Logic) × 2(Conflict/No-conflict) × 2(MP/Disjunctives) × 2(Experimental/ Control Condition) mixed design ANOVA was used for the second set of analyses measuring the effects under belief and logic instruction (Table 6). As in Experiment 1, we conducted both a subjects and items analysis on accuracy scores and present both F scores below.

The results produced a main effect of *instruction* for subjects (Belief/ Logic), $F_1(1, 71) = 22.482$, $p < .001$, $\eta_p^2 = 0.240$; and items, $F_2(1, 120) = 13.783$, $p < .001$, $\eta_p^2 = 0.103$, with better performance on logic-based judgments than belief judgments *(92% logic vs. 89% belief)*. There was a main effect of *problem* type for subjects (Conflict/No-conflict), $F_1(1, 71) = 37.058$, $p < .001$, $\eta_p^2 = 0.343$; and items, $F_2(1, 120) = 30.573$, $p < .001$, $\eta_p^2 = .203$, with conflict items producing lower accuracy scores than no-conflict items *(87% vs. 93%)*. There was also a main effect of *complexity* for subjects (Disjunctives/MP), $F_1(1, 71) = 37.944$, $p < .001$, $\eta_p^2 = 0.348$; and items, $F_2(1, 120) = 16.613$, $p < .001$, $\eta_p^2 = 0.122$, with poorer performance on disjunctives than MP judgments *(88% vs. 92%)* but no main effect of *condition* for subjects analysis, $F_1(2, 71) = 0.909$, $p = .344$, $\eta_p^2 = 0.013$; signifying no difference in accuracy scores across conditions *(91% experimental vs.*

89% control), although this effect was marked as significant in the items analysis, $F_2(1, 120) = 8.841$, $p = .004$, $\eta_p^2 = 0.066$.

There was a significant interaction between *instruction* and *problem* type for the subject's analysis, $F_1(1, 71) = 16.034$, $p < .001$, $\eta_p^2 = .184$; and items analysis, $F_2(1, 120) = 13.900$, $p < .001$, $\eta_p^2 = .104$, showing, in line with Experiment 1, that conflict has less of an impact on conclusion validity *(91% L-conflict vs. 93% L-no-conflict)* compared to the believability of a conclusion *(84% B-conflict vs. 94% B-no-conflict* Table 6). There was a significant interaction between *problem* type and *complexity* for subjects, $F_1(1, 71) = 10.255$, $p = .002$, $\eta_p^2 = 0.126$; and items, $F_2(1, 120) = 5.104$, $p = .026$, $\eta_p^2 = 0.041$, with conflict having less of an impact on disjunctive judgments *(87% Conflict vs. 90% No-conflict)* than MP judgments *(88% Conflict vs. 97% No-conflict)*. There was also an interaction between *instruction* and *complexity* for subjects, $F_1(1, 71) = 27.773$, $p < .001$, $\eta_p^2 = 0.281$; and items, $F_2(1, 120) = 9.200$, $p = .003$, $\eta_p^2 = 0.017$, with a bigger difference in performance between disjunctives and MP judgments under logic instruction *(89% L-Disjunctives vs. 96% L-MP)* than under belief instruction *(88% B-Disjunctives vs. 89% B-MP)*. Finally, there was a 3-way interaction between *instruction, problem* type and *condition* for subjects, $F_1(1, 71) = 9.035$, $p = .004$, $\eta_p^2 = 0.113$; and items, $F_2(1, 120) = 18.472$, $p < .001$, $\eta_p^2 = 0.133$ as demonstrated in Figure 4.

As in Experiment 1, a follow-up analysis on both conditions confirmed that the *instruction* by *problem* type interaction was present in the experimental condition for subjects, $F_1(1, 36) = 18.727$, $p < .001$, $\eta_p^2 = 0.342$; and items, $F_2(1, 120) = 30.763$, $p < .001$, $\eta_p^2 = 0.204$, but absent in the control condition for both subjects, $F_1(1, 35) = 0.745$, $p = .394$, $\eta_p^2 = 0.021$; and items, $F_2(1, 120) = 1.044$, $p = .309$, $\eta_p^2 = 0.009$. This demonstrates that the effect of conflict was greater in the experimental condition on belief judgments *(82% B-Conflict vs. 95% B-No-conflict)* compared to logic judgments *(94% L-Conflict vs. 94% L-No-conflict),* whereas in the control condition, the effect of conflict was similar under both belief (*85% B-Conflict vs. 92% B-No-conflict)* and logic instruction *(89% L-Conflict vs. 92% L-No-conflict)* (Table 6).

A mixed design ANOVA on response latencies replicated most of the main effects in Experiment 1. Results showed no main effect of *instruction* for subjects, $F_1(1, 71) = 0.796$, $p = .375$, $\eta_p^2 = 0.011$; or items, $F_2(1, 120) = .258$, $p = .612$, $\eta_p^2 = 0.002$, indicating no significant difference in response latencies for belief *(6195ms)* and logic *(6320ms)* judgments, but there was a main effect of *problem* type for subjects, $F_1(1, 71) = 9.097$, $p = .004$, $\eta_p^2 = 0.114$; with a marginal effect in the items analysis, $F_2(1, 120) = 3.167$, $p = .078$, $\eta_p^2 = 0.026$, showing conflict items taking longer than no-conflict items (*6428ms vs. 6086ms*). There was also a main effect of *complexity* for both subjects, $F_1(1, 71) = 6.009$, $p = .017$, $\eta_p^2 = 0.078$; and items, F_2(1,

120) = 6.140, $p = .015$, $\eta_p^2 = 0.049$, demonstrating quicker response times to MP arguments compared to disjunctive arguments (*6141ms vs. 6396ms*), but there was no main effect of *condition* in the subjects analysis, $F_1(2, 71) = 1.217$, $p = .274$, $\eta_p^2 = 0.017$; whereas the items analysis did produce a significant effect, $F_2(1, 120) = 16.601$, $p < .001$, $\eta_p^2 = 0.122$, with slower latencies in the experimental condition (*6,452ms*) compared to the control condition (*6,085ms*).

There was a marginal interaction between *instruction* and *condition* for subjects, $F_1(1, 71) = 3.949$, $p = .051$, $\eta_p^2 = 0.053$; but no interaction in the item's analysis, $F_2(1, 120) = 0.060$, $p = .807$, $\eta_p^2 < 0.001$. There was a significant interaction between *problem* type and *complexity* for subjects, $F_1(1, 71) = 4.161$, $p = .045$, $\eta_p^2 = 0.055$; which again was not significant in the item's analysis, $F_2(1, 120) = 0.146$, $p = .703$, $\eta_p^2 = 0.001$. Finally, an interaction between *instruction* and *complexity* for subjects, $F_1(1, 71) = 18.414$, $p <. 001$, $\eta_p^2 = 0.206$; and items, $F_2(1, 120) = 4.270$, $p = .041$, $\eta_p^2 = 0.034$, showed a bigger difference in response times between MP and disjunctive judgments under logic instruction *(6027ms L-MP vs. 6613ms L-Disjunctives)* than under belief instruction *(6256ms B-MP vs. 6179ms B-Disjunctives).*

Discussion

The principle objective of Experiment 2 was to examine whether the effect of the Stroop instruction on the accuracy of belief judgments in Experiment 1 could be explained in terms of (1) the **inhibitory demands** of the Stroop instruction or (2) the **additional demands** of including a third instructional cue. This was achieved by changing the instructional condition from one that required the inhibition of word naming to one that demanded memory updating. The memory updating instruction asked participants to remember the font characteristics of the concluding word of the previous trial and make a judgment on whether it matched the font in the current trial. Given that participants were unaware of the specific instruction from trial to trial, accuracy depended upon continuously updating a memory of the previous trial font. Consistent with explanation (2), Experiment 2 showed that logic had its biggest impact on belief judgments when an additional, cognitively demanding, instructional condition was introduced. The impact was larger in the Experimental condition, compared to the simple matching judgment required in the Control condition. This suggests that the increased executive demands created by the extra response alternative made it more challenging to generate a logical output.

We would argue that in both experiments 1 and 2, the presence of a third demanding instructional cue increases cognitive demands by requiring participants to inhibit two competing responses in favour of a target response on each trial. The finding suggests that it is not the specific requirements of the third instruction (Stroop vs. n-back) that results in interference, but the simple presence of competing response options. Again, this substantiates a parallel processing model where logical inferences are available early and require inhibiting in order to allow the more effortful belief response to complete.

The analysis of belief and logic instruction replicated the main effects shown in Experiment 1, with belief judgments and conflict problems being more prone to errors, and more complex disjunctive arguments producing lower accuracy scores. Furthermore, the presence of a conflict had a larger effect on the accuracy of belief judgments compared to logic judgments, supporting previous claims that belief judgments are effortful and draw on inhibitory resources (Howarth et al. 2016).

Finally, although participants were more accurate in the control condition, they also took significantly longer to respond. One possible explanation of this finding is that in the *n*-back condition participants were required to consistently keep an active representation of the last word in mind from trial to trial, whereas with the matching task there was no active representation. Instead participants were required to re-inspect the first premise in order to make a match with the last word in the concluding statement. This would have required an additional saccade and therefore took longer to complete.

General discussion

In this article, we examined the impact of introducing a third competing instructional condition on participants' ability to accurately judge the validity or the believability of a conclusion. Previous research has shown that belief-logic conflict has a greater influence on belief judgments than logic judgments (Handley et al., 2011; Howarth et al., 2016). One interpretation of this finding is that a logical response is available early and must be inhibited in order to generate a slower response based upon underlying beliefs. The presence of a third cognitively demanding instructional condition was predicted to impact more upon effortful belief judgments than logic judgments. This effect was present in both experiments irrespective of whether the executive demands of the additional instruction depended upon the inhibition of word naming or the updating of working memory.

Our main focus was to explore the function of inhibition on reasoning under conflict in relation to the Handley and Trippas (2015) parallel processing model (Figure 1). The findings lend support to this model and suggest that belief judgments on this task are more effortful than logic judgments; hence the logical status of a conclusion interferes more with belief responses under conflict. The implication is that some logic-based reasoning resembles an intuitive type-1 process, and according to the model must be inhibited in favour of a more deliberative belief-based output. Introducing a third instructional condition, in the form of colour-naming or memory updating, impacted on the ability to inhibit the early completion of a logical response. In other words, increasing the demands on executive resources enhanced the impact of logic on effortful belief judgments by burdening the resources required to inhibit intuitive logic. Importantly, the nature of this additional instructional cue had little differential effect on the overall impact of conflict.

There was, however, an inconsistency between the accuracy and latency data. While the accuracy data support the idea that belief judgments require effortful processing, the latency data suggests that increasing the number of competing responses has a bigger impact on the time it takes to reason on the basis of logic (Experiment 2) specifically logic-conflict items (Experiment 1) more so than those under belief instruction. We would argue that the discrepancy between the accuracy and latency data corroborates there being qualitative differences between logical outputs. In line with recent work (Bago & De Neys, 2017; Newman et al., 2017), we would argue that there are "two routes" to a logical solution; an intuitive route and a deliberative route. The intuitive route is a type-1 process, which completes outside explicit awareness but is sensitive to the logical structure of a problem. This automatic logical output is probably accompanied by a Feeling of Rightness (Thompson, Prowse, Turner, & Pennycook, 2011) based on conceptual fluency (Morsanyi & Handley, 2012), or intuitive strength (Bago & De Neys, 2017), which provides the reasoner with an intuitive logical cue. The impact of a conflicting logical response on belief judgments may be interpreted as a type-1/type-2 conflict requiring the inhibition of this intuitive logical cue in order to reason according to beliefs. The cue, however, may not be sufficient to generate a logical response. When instructed to reason logically, a second deliberative route must be engaged, which involves effortful type-2 processing that depends upon the explicit integration of premise information and model construction, triggered in parallel to a belief-based output. This would explain the impact of a secondary task on logic judgments reported by Howarth et al. (2016), where extracting the underlying structure of an argument when instructed to reason logically, was impacted on by requirement to generate random numbers (Random Number Generation task) alongside the reasoning task. When additional response alternatives make the task more cognitively demanding (Experiments 1 and 2) inhibition of the initial logical output becomes more challenging when instructed to reason on the basis of beliefs, and more time consuming when integrating premise information under logic instruction. However, due to the relatively simple nature of the conditional problems the presence of an additional instructional condition has less impact on logic accuracy.

Recent work by Newman, Gibb and Thompson (2017) using challenging response deadlines provides evidence consistent with this account. They showed evidence of sensitivity to logical structure under response deadline and under free responding consistent with a dual route to logical judgments (rule-based). Interestingly, they also showed sensitivity to counterexample availability under speeded task suggesting that associative beliefs were available through type-1 processing, a finding consistent with dual route accounts of reasoning (Markovits, Brisson, de Chantal, & Thompson, 2017). Their findings suggest that reasoners are intuitively sensitive to logical and probabilistic information (Route 1) but when given more time to reflect on their responses, the ability to discriminate between valid and invalid inferences improves, as evidence for slower, rule-based processing (Route 2).

Our findings are consistent with dual routes to a logical solution, but we would argue that there is limited evidence in our data to support dual routes to a belief-based response. On the contrary, the uni-directional impact of logic on belief judgments, particularly under conditions of additional executive demands, suggests that in this paradigm beliefs operate at a type-2 level. In both Experiments 1 and 2, there was no effect of beliefs on logic judgments in the experimental conditions and if anything, logical performance on conflict problems was marginally improved. This observation is consistent with evidence of improved logical performance in other paradigms, under cognitive load, such as scalar implicature (De Neys & Schaeken, 2007), or most recently shown with base rate and syllogistic reasoning problems where logical responding increases under load and response deadline (Bago & De-Neys, 2017).

Is it possible that the belief judgments in these studies are more difficult because of the way in which the problems were structured? Consider for example one of the MP problems:

If the bird is a Dove then it is orange

Suppose the bird is a Dove

Does it follow that the bird is orange?

As a reviewer pointed out, one cannot make a judgment concerning the believability of the conclusion with reference to the conclusion alone. In order to determine its believability, one also needs to know that the bird being referred to is a Dove, which can only be determined with reference to the second premise also. So, like logical judgments, there is some integration of premise information required to generate a response. We do not think that this problem feature can uniquely explain why belief judgments are more effortful on this task than logic judgments for three reasons. Firstly, whilst belief judgments on these problem types require reference to one premise and a conclusion, logic judgments require significant additional processing through the integration of both presented premises. Therefore, on the basis of premise integration, one would expect belief judgments to be simpler. Secondly, if we consider the disjunctive arguments employed (Tables 1 and 4), in all cases the belief judgment can be made with reference only to the concluding sentence. There is no evidence that the greater difficulty of belief judgments is confined to the MP arguments. Thirdly, there is evidence from previous research (Handley et al., 2011; Study 4) that logic impacts upon belief judgments on simple MP problems, where conclusion believability is manipulated entirely independent of belief in the major premise through the use of abstract middle terms.

Our findings fit well with a parallel competitive dual process model, however could a continuous rather than dichotomous scale of complexity account for our data? Keren and Schul's (2009) uni-model offers a single evaluative system to distinguish between the structure of the argument or knowledge of its content as distinct external criteria and alternative responses come in and out of awareness continuously when reasoning. This shift between mental states to solve different problems, for example, will depend on features such level of control, awareness and time. The way these features join depends on the goal and the environmental limitations. Similarly, Kruglanski and Gigerenzer (2011) reject the dual process claims of qualitatively distinct processing systems underlying responses to reasoning and judgment tasks. Instead they suggest that both intuitive and deliberative judgments rely on the application of rules that are activated by task features and environmental cues. The nature of the rules that are applied will be determined by memory capacity, attentional demands and ecological fit.

In some ways, Keren and Schul's uni-model has aspects in common with the parallel competitive theory, in that it allows for the parallel activation of competing cues to a problem solution. However, neither this account nor Kruglanski and Gigerenzer's position supports a qualitative difference between type-1 and type-2 responses to reasoning tasks. We would argue that the data in this article and elsewhere provide strong evidence, at the very least, of qualitatively distinct processes underlying logical responding to reasoning tasks. As discussed earlier, the findings here show an informative dissociation between the accuracy and latency data, suggesting an intuitive sensitivity to logical structure which impacts on the accuracy of belief judgments, coupled with more deliberative and time dependent responding under logical instruction. While this is not decisive evidence of a qualitative distinction, other findings suggest a clear dissociation between explicit and intuitive logical sensitivity. For example, Morsanyi and Handley (2012) showed that participants "liked" sentences that logically followed from the sentences that preceded them more than those that did not. Crucially this effect was uninfluenced by participants cognitive capacity, mindset or superficial problem structures, features that did influence explicit reasoning judgments. We have recently shown that validity not only influences how much participants like a presented statement, but also their perceptual judgments concerning its brightness, a feature completely disconnected from its logical validity (Trippas, Handley, Verde & Morsanyi, 2016). It has been suggested that this effect arises because of a feeling of fluency arising from reading a series of logically coherent sentences that in turn generates a feeling of positive affect which is then misattributed to the perceptual features of the stimuli. These findings, coupled with evidence of early detection of logical validity under rapid responding conditions (Bago & De Neys, 2017; Newman et al., 2017), strongly suggest that there are qualitatively distinct processes underlying sensitivity to logical structure.

Conclusion

With respect to traditional Default Interventionist (DI) dual process accounts of reasoning (see Evans & Stanovich, 2013), the data presented here does not fit well for two key reasons; (1) The findings strongly suggest type-2 processing underpins belief-based judgments and (2) simple arguments appear to cue an intuitive logical response, perhaps accompanied by a feeling of rightness associated with the validity of the argument. Thus, our data fit better with a model where both the logical structure of an argument and relevant knowledge are processed simultaneously and conflict is explained in terms of the initiation of competing cues to a solution. The direction of conflict, however, will depend on the complexity of the problem, the instructions delivered and the ability to draw on the suitable executive resources for type-2 processing. With simple logical arguments of the kind used here, an intuitive logical response is available early and must be inhibited in order to explicitly generate a response based on belief. Withholding an available logical response is effortful and, as our findings show, increasing cognitive demands significantly reduces participants' capacity to do so. In contrast, as previous work has shown, on more complex arguments such as syllogisms, a belief-based response may complete first, hence impacting on logic-based reasoning (Trippas, Thompson & Handley, 2017).

While the evidence of dual routes to a logical solution is consistent with the parallel processing model, the model also makes clear predictions concerning the impact of complexity, not only on logic judgments, but also on judgments of belief. Future research could usefully focus on the manipulation of the complexity of belief judgment by using materials that vary in terms of degree of belief and examine the impact on logic-based reasoning as well as measuring how logical processing compromises beliefs. One might expect that with beliefs that are more complex, to evaluate the impact of conflicting logical structure will be greater. Furthermore, research on the development of inhibition and its impact on reasoning could be valuable. To date research on inhibition has shown that its capacity produces a curvilinear age trend. The ability to resist prepotent responses tends to improve from childhood to adolescence and declines again in later life (Bedard et al., 2002; Christ, White, Mandernach & Keys, 2001). De Neys and Van Geldor (2009) demonstrated with syllogistic reasoning, that when belief and logic conflict, performance is determined by a person's aptitude for inhibiting a response. Therefore, one would expect no effect of conflict if inhibition resources were underdeveloped (in young children) or limited in any way.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

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

Bedard, A.-C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, *21*(1), 93–111.

Christ, S. E., White, D. A., Mandernach, T., & Keys, B. A. (2001). Inhibitory control across the life span. *Developmental Neuropsychology*, *20*(3), 653–669.

De Neys, W. (2006). Dual Processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*(5), 428–433.

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

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS One*, *6*(1), e15954. doi: 15910.11371/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.

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

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, and Behavioral Neuroscience*, *10*(2), 208–216.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133.

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: the rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science*, *12*(1), 123–130.

Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.

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

Evans, J. St. B. T. (2010a). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.

Evans, J. St. B. T. (2010b). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, *21*(4), 313–326.

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

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. doi:10.1111/j.1467-9280.2006.01681.x

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis*. 6th ed. Uppersaddle River: Pearson Prentice Hall.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking and Reasoning*, *10*(2), 175–195. doi:10.1080/13546780442000051

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), 28–34.

Handley, S. J., & Trippas, D. (2015). Chapter two-dual processes and the interplay between knowledge and structure: A new parallel processing model. *Psychology of Learning and Motivation*, *62*, 33–58.

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), 330–349.

Kahneman, D. (2011). *Thinking, fast and slow*: New York: Farrar, Straus and Giroux.

Keren, G., & Schul, Y. (2009). Two is not always better than one a critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

Kruglanski, Arie W. & Gigerenzer, Gerd. (2011). "Intuitive and deliberate judgments are based on common principles": Correction to Kruglanski and Gigerenzer. *Psychological Review*, *118*(3), 522–522.

Markovits, H., Brisson, J., de Chantal, P. L., & Thompson, V. A. (2017). Interactions between inferential strategies and belief bias. *Memory and Cognition*, *45*(7), 1182–1192.

Mat Roni, S. (2014). Introduction to SPSS, School of Business, Edith Cowan University, Australia. Retrieved from [2016.10. 14] from www.researchgate.net/publication/262151892_Introduction_to_SPSS.

Milligan, G. W. (1987). The use of the arc-sine transformation in the analysis of variance. *Educational and Psychological Measurement*, *47*(3), 563–573.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*(2), 111–121.

Morsanyi, K., & Handley, S. J. (2012). Logic feels so good—I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 596–616.

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.

Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counter-examples: Deductive rationality in human reasoning. *The Quarterly Journal of Experimental Psychology: Section A*, *56*(7), 1129–1145.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Stupple, E. J., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking and Reasoning*, *14*(2), 168–181.

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

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition*, *39*(7), 1275–1289.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking and Reasoning*, *20*(2), 147–168.

Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1448.

Trippas, D., Thompson, V. A., & Handley, S. J. (2016). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory and Cognition*, 1–14.

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 & Cognition*, *45*(4), 539–552.