



PEARL

**Using forced choice to test belief bias in syllogistic reasoning.**

Trippas, Dries; Verde, Michael F.; Handley, Simon J.

**Published in:**  
Cognition

**DOI:**  
[10.1016/j.cognition.2014.08.009](https://doi.org/10.1016/j.cognition.2014.08.009)

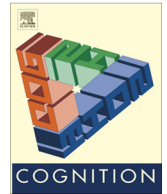
**Publication date:**  
2014

**Link:**  
[Link to publication in PEARL](#)

**Citation for published version (APA):**

Trippas, D., Verde, M. F., & Handley, S. J. (2014). Using forced choice to test belief bias in syllogistic reasoning. *Cognition*, 133(3), 586-600.  
<https://doi.org/10.1016/j.cognition.2014.08.009>

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Wherever possible please cite the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.



# Using forced choice to test belief bias in syllogistic reasoning



Dries Trippas\*, Michael F. Verde, Simon J. Handley

Plymouth University, UK

## ARTICLE INFO

### Article history:

Received 6 November 2013

Revised 12 August 2014

Accepted 14 August 2014

Available online 18 September 2014

### Keywords:

Belief bias

Signal detection theory

Forced choice

Individual differences

## ABSTRACT

In deductive reasoning, believable conclusions are more likely to be accepted regardless of their validity. Although many theories argue that this *belief bias* reflects a change in the quality of reasoning, distinguishing qualitative changes from simple response biases can be difficult (Dube, Rotello, & Heit, 2010). We introduced a novel procedure that controls for response bias. In Experiments 1 and 2, the task required judging which of two simultaneously presented syllogisms was valid. Surprisingly, there was no evidence for belief bias with this forced choice procedure. In Experiment 3, the procedure was modified so that only one set of premises was viewable at a time. An effect of beliefs emerged: unbelievable conclusions were judged more accurately, supporting the claim that beliefs affect the quality of reasoning. Experiments 4 and 5 replicated and extended this finding, showing that the effect was mediated by individual differences in cognitive ability and analytic cognitive style. Although the positive findings of Experiments 3–5 are most relevant to the debate about the mechanisms underlying belief bias, the null findings of Experiments 1 and 2 offer insight into how the presentation of an argument influences the manner in which people reason.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

An argument is logically valid if its conclusion necessarily follows from its premises. From a normative standpoint, the logical status of an argument is determined exclusively by its structure. Nevertheless, it is well known that other factors routinely influence peoples' judgments of argument validity. One of the most important of these is the extent to which the conclusion of an argument conforms with a person's pre-existing view of the world. People tend to be easily persuaded by invalid arguments that fit their beliefs yet are likely to resist valid arguments that oppose their beliefs. This phenomenon is known as belief bias (Wilkins, 1928).

Numerous psychological theories have been developed to account for belief bias in syllogistic reasoning (e.g., Evans, Barston, & Pollard, 1983; Quayle & Ball, 2000; Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003). Mental model theory (MMT; Oakhill, Johnson-Laird, & Garnham, 1989) holds that people construct mental models to simulate the components of an argument. When faced with unbelievable conclusions, they engage in a more thorough search for alternative models, improving the likelihood of arriving at the normatively correct conclusion. Selective processing theory (SPT; Evans, Handley, & Harper, 2001; Klauer, Musch, & Naumer, 2000) posits that reasoning strategy depends on conclusion believability: people attempt to confirm believable conclusions but disconfirm unbelievable ones. According to dual process theories like those described by Evans (2007), unbelievable conclusions may recruit additional, reflective processing. All of these theories suggest that conclusion believability can influence the quality of reasoning. However, it is also

\* Corresponding author. Address: Cognition Institute, School of Psychology, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK. Tel.: +44 7801 518311.

E-mail address: [dries.trippas@plymouth.ac.uk](mailto:dries.trippas@plymouth.ac.uk) (D. Trippas).

generally accepted that believability may exert an influence independent of reasoning quality in the form of a heuristic tendency to accept believable and reject unbelievable conclusions. Such a tendency is a response bias that affects the pattern of validity judgments without altering their overall accuracy. Of course, a simple alternative to the theories described above is that the effect of beliefs is purely one of response bias. Recent work by [Dube, Rotello, and Heit \(2010\)](#) suggests that this alternative view of belief bias is not only viable but one that traditional methods of analysis make difficult to rule out.

In a typical belief bias experiment, participants judge the logical validity of conclusions that vary in their congruence with prior belief. Analysis has traditionally focused on three aspects of these endorsement rates (e.g., [Evans & Curtis-Holmes, 2005](#)). The main effect of logic, the tendency to endorse valid over invalid conclusions, indicates deductive reasoning capacity. The main effect of belief, the tendency to endorse believable over unbelievable conclusions, is thought to be driven by a belief-based heuristic. Finally, a logic  $\times$  belief interaction is taken as evidence for a qualitative effect of beliefs on reasoning that goes beyond response bias. Although it might seem intuitively reasonable, this manner of interpreting the data has been the subject of recent critical attention. In particular, the conclusion that a logic  $\times$  belief interaction necessarily signifies a change in the quality of reasoning requires assumptions about the nature of the underlying processes that may not be justified ([Dube et al., 2010](#); [Klauer et al., 2000](#)). This is a serious issue given that much of the theoretical debate surrounding belief bias has centered on evidence for this interaction.

Are there alternative ways of determining whether a change in the pattern of endorsement rates is the result of a change in the quality of reasoning, as the major theories of belief bias suggest, rather than a change in response bias? Although making the distinction is theoretically critical, it is not straightforward. Some recent studies of syllogistic reasoning have turned to signal detection theory (SDT; [Macmillan & Creelman, 2005](#)) in order to model the effect of beliefs on accuracy and response bias ([Dube et al., 2010](#); [Heit & Rotello, 2014](#); [Trippas, Handley, & Verde, 2013, 2014](#)). In the SDT model, the reasoning process yields evidence of validity that is represented as a continuous strength variable. Valid and invalid arguments are described by distributions of evidence strength. People judge validity by setting a criterion which evidence must exceed if an argument is to be endorsed as “valid.” The endorsement rates are determined by the proportions of the distributions that exceed the criterion. Accuracy depends on the overlap of the evidence distributions. When a change in the reasoning process brings about separation of the distributions, this alters the pattern of endorsement rates and, more importantly, improves accuracy. A shift in response bias that makes the criterion more liberal or conservative also alters the pattern of endorsement rates but will not affect accuracy. An effect on judgment accuracy indicates that something other than response bias is at work.

A commonly used SDT tool for examining movement in evidence distributions is the receiver operating

characteristic (ROC) which plots the ratio of hits and false alarms at several levels of confidence, allowing inferences to be made about the distributions. [Dube et al.’s \(2010\)](#) analysis of ROCs in syllogistic reasoning showed no effect of conclusion believability on accuracy, a result consistent with a pure response bias account of belief bias. [Dube et al.](#) also pointed out that the shape of empirical ROCs argues against the validity of drawing conclusions about accuracy based on a logic  $\times$  belief interaction in endorsement rates. According to the SDT model, such an interaction could come about entirely as a result of response bias. In a follow-up study using similar methods, [Trippas et al. \(2013\)](#) partly replicated their findings but found that under some conditions, such as with higher cognitive ability and relatively complex problem types, effects of beliefs on reasoning quality did emerge.

The ROC findings offer somewhat inconsistent support for theories of belief bias that posit qualitative effects on reasoning. The ROC method also has several potential shortcomings. One is practicality. The method requires capturing performance over a range of criterion levels. This can be achieved by collecting confidence ratings alongside validity judgments or by manipulating response bias across multiple groups. Either method is data-intensive and time consuming, and the subsequent analysis is more complex than dealing with a single set of endorsement rates. Collecting confidence ratings is the more popular of the two methods because it allows a within-subjects treatment. However, the source of confidence in syllogistic reasoning has not been extensively studied. There is evidence for alignment between normative responding and confidence-type judgements (e.g., [De Neys, 2012](#); [Morsanyi & Handley, 2012](#)), but confidence may not always be related to accuracy ([Shynkaruk & Thompson, 2006](#)). Finally, there is ongoing debate about whether SDT provides the optimum model for empirical ROCs ([Dube, Rotello, & Heit, 2011](#); [Klauer & Kellen, 2011](#)).

### 1.1. The forced choice method

Forced choice is a procedure commonly used in domains such as perception and memory. Although it has not to our knowledge been applied to the study of syllogistic reasoning, its properties make it ideal for investigating belief bias. In the traditional single-probe procedure, participants are presented with individual syllogisms and must decide whether the conclusion is valid or invalid. In the forced choice procedure which we introduce here, two syllogisms are presented side by side and the participant must decide which of the two has the valid conclusion. The procedure is not much more complex than the traditional one, and the dependent measure, proportion correct, is simple and straightforward to interpret. Most importantly, forced choice offers a means of isolating changes in reasoning from changes in response bias by removing the latter from the decision process. This is done by equating the believability of competing conclusions.<sup>1</sup>

<sup>1</sup> The procedure does allow one form of bias: participants could show an arbitrary preference for the right or left display position. Spatial position bias was examined in our analysis.

**Table 1**  
Examples of the believable and unbelievable problem types used in Experiment 1.

Problem type	Valid syllogism	Invalid syllogism
Believable	Some animals are piffures	Some dogs are piffures
	No piffures are dogs Some animals are not dogs	No piffures are animals Some animals are not dogs
Unbelievable	Some dogs are piffures	Some animals are piffures
	No piffures are animals Some dogs are not animals	No piffures are dogs Some dogs are not animals

Note: In the actual Experiments all four figures and both conclusion directions were used. The valid argument was presented on the left half the time and on the right half the time.

Consider the example problems shown in Table 1. In the believable condition, the two simultaneously presented conclusions are both believable. In the unbelievable condition, the conclusions are both unbelievable. Prior beliefs no longer provide a clear and easy basis for judging validity. Because a preference for endorsing believable conclusions can have no differential effect on believable and unbelievable trials, any difference in accuracy between the conditions can be attributed to the influence of conclusion believability on the way that people reason.

In the experiments to follow, the focus will be on the empirical question of whether beliefs affect validity judgments in the absence of response bias. Because the logic of excluding response bias with forced choice is theoretically neutral, the method also offers a way to provide converging support for the conclusion drawn from ROC analysis that unbelievable arguments encourage better reasoning in a manner mediated by cognitive ability (Trippas et al., 2013).

## 2. Experiment 1

We tailored the forced choice method to syllogistic reasoning as follows. Each trial consisted of two simultaneously presented syllogisms, one of which was logically valid and the other logically invalid. Participants were instructed to reason about both syllogisms and to choose the valid one. The influence of beliefs on the reasoning process was examined by comparing the accuracy of judgments in two conditions. In the believable condition, both arguments had believable conclusions (top row of Table 1). In the unbelievable condition, both arguments had unbelievable conclusions (bottom row of Table 1). If belief bias is no more than a heuristic tendency to prefer believable over unbelievable conclusions, accuracy should be the same in both conditions. On the other hand, if people reason more effectively when confronted with unbelievable conclusions as some theories suggest, accuracy should be higher in the unbelievable conclusion condition. In addition to the believable and unbelievable conditions, a neutral condition in which end terms were replaced with nonsense words was added to serve as a baseline.

## 2.1. Method

### 2.1.1. Participants

A total of 22 participants took part in exchange for course credit. Seven participants were male and 15 were female (age 18–31,  $M = 21$ ,  $SD = 3$ ).

### 2.1.2. Materials

A list containing 48 problems was created to be used as stimuli. Every problem consisted of two syllogisms presented side by side. The two syllogisms always had identical conclusions. One syllogism was always valid and the other one was always invalid. The valid argument was presented equally often in the left and right positions. Sixteen argument structures taken from Dube et al. (2010) were repeated three times. Half of these were valid and half were (indeterminately) invalid. All syllogisms were complex, meaning that multiple representations of the premises could be constructed. Potential unwanted effects of problem content were controlled for by randomly assigning the problem contents to the syllogisms for each participant anew (Trippas et al., 2013). This resulted in a unique problem list for every participant. Conclusion believability was manipulated by combining sixteen superordinate categories (e.g., birds) with two subordinate members of each those categories (e.g., parrots and sparrows; see Appendix A for the full list).

Believable, unbelievable and neutral problem types were created to be used as stimuli. Believable problem trials featured a valid–believable syllogism beside an invalid–believable one (see the top panel of Table 1 for an example). Believable conclusions were created by presenting the superordinate category before a subordinate member of that category (e.g., some birds are not parrots). Unbelievable problem trials featured a valid–unbelievable syllogism next to an invalid–unbelievable one. Unbelievable conclusions were created by reversing the assignment order used in the believable case (e.g., some parrots are not birds). We controlled for premise believability using nonsense words for middle terms (Newstead, Pollard, Evans, & Allen, 1992). Forty-eight pseudowords (Appendix A) were generated (Keuleers & Brysbaert, 2010) and randomly assigned to the middle terms of the syllogisms. Neutral problems consisted of a valid–neutral syllogism next to an invalid–neutral one. Neutral conclusions were created by presenting two pseudowords (e.g., some fromps are not blarks). The middle-term of the premises of neutral problems was always assigned a randomly chosen subordinate member that was unused in the believable and unbelievable problem types. Combining these syllogistic structures and item contents yielded a uniquely randomised list of 48 problems (16 of each type) for every participant.

### 2.1.3. Procedure

The participants were tested on individual computers in small groups. Upon entering the lab, they were randomly assigned to a computer and were told to read an instruction sheet which briefly explained the nature of the task. The sheet contained a definition of logical validity which stressed that participants should assume the truth of the

premises and to only consider an argument valid if its conclusion necessarily followed from this premises. An example of a valid and an invalid syllogism was provided. Further instructions were presented on the screen:

*“In the following experiment you will repeatedly be presented with two reasoning problems. One of these problems is logically valid. One of these problems is logically invalid. It is your task to choose which problem is valid. Use the mouse to click the box that contains the valid argument according to you. The box around the reasoning problem you chose will turn green, indicating that you think it is valid. The box around the reasoning you didn’t choose will turn red, indicating that you think it is invalid. After every choice, use your mouse to indicate how confident you are that you made the correct decision. Before the actual experiment you will be presented with 6 practice trials to ensure that you understand the instructions. After the practice trials the actual experiment begins. If you have any questions ask the experimenter before or during the practice trials.”*

The participants then completed six practice trials (two believable, two unbelievable and two neutral). No accuracy feedback was provided. If the participants had no further questions after completing the practice trials, they were told to complete the 48 experimental trials at their own pace. Problems were presented to each participant in a uniquely randomised order. For every problem, two syllogisms were shown side by side. Both syllogisms had identical conclusions and item contents. When one of the response options was chosen by clicking on it, the border of the box containing the chosen syllogism turned green and the word “VALID” appeared underneath. Simultaneously, the border of the box containing the remaining syllogism turned red and the word “INVALID” appeared underneath. Participants were allowed to change their selection as often as they wished by clicking the other box, although they rarely did so (<2% changes). After the experiment participants were thanked and debriefed.

#### 2.1.4. Measures and design

The participant’s choice of argument (left vs. right) and the accuracy of the choice (whether the valid response was chosen) was recorded on every trial. The experiment had a one-way design with problem type (believable vs. unbelievable vs. neutral) manipulated within subjects.

## 2.2. Results

### 2.2.1. Data treatment

Given the binary nature of our outcome variables, we used generalized linear mixed models (GLMM) with a logit link, binomially distributed residuals, and contrast coding to analyze our data (e.g., Barr, Levy, Scheepers, & Tily, 2013; Jaeger, 2008). Models were fit using the `glmer()` function from the `lme4` package (Bates, Maechler, & Bolker, 2013; implemented in R statistics, R Core Team, 2012). Formal model specifications for each reported analysis can be found in Appendix B. We tested for significance using likelihood ratio tests (LRT; e.g., Bates, 2010) where possible and Wald tests otherwise. LRTs consist of

comparisons between full models and nested, more restricted models. The difference in fit between both models (deviance) is then compared with a  $\chi^2$  distribution with degrees of freedom equal to the amount of constrained parameters to see if the imposed restriction significantly reduces the fit of the model. Significant reductions in fit indicate that constraining the parameters to be equal is not afforded by the data (i.e., that there is an effect). Tests for spatial bias and above chance performance used Wald tests because LRTs are infeasible for intercept-only models. All results are reported in terms of both log-odds (LO) and proportions correct. These marginal means were calculated by transforming the LO parameter estimates into a probabilities using the inverse logit function.

### 2.2.2. Preliminary analyses

To rule out bias based on spatial position, we fit an intercept only logit GLMM with left selections as the dependent variable (dummy coded 1 = left, 0 = right). There was no preference for the left or the right argument,  $LO = 0.08$ ,  $z = 1.13$ ,  $p = .26$ . The overall probability of choosing the argument in the left position was 52%.

We also tested whether performance was above chance overall using an intercept only logit GLMM with correct responses as the dependent variable (1 = correct, 0 = incorrect). Participants responded significantly above chance,  $LO = 0.76$ ,  $z = 5.40$ ,  $p < .001$ . Overall proportion correct was 68%.

### 2.2.3. Believable vs. unbelievable

A summary of performance can be found in Table 2. The critical test for an effect of beliefs on reasoning was the comparison of the believable and unbelievable conditions. We analyzed correct responses (1 = correct, 0 = incorrect) as a function of problem type (believable vs. unbelievable) using a logit GLMM. There was no significant difference in proportion correct between the believable and unbelievable problems,  $\chi^2(1) = 0.19$ ,  $p = .66$ .

A neutral problem type was also included as a baseline condition against which to compare the believable and unbelievable problem types. Logit GLMM regression of problem type (believable vs. unbelievable vs. neutral) on correct responses revealed that there was no effect of problem type,  $\chi^2(2) = 0.65$ ,  $p = .72$ .

## 2.3. Discussion

If unbelievable conclusions encourage people to reason more effectively, participants should have been more accurate on unbelievable trials. We observed no significant

**Table 2**  
Log Odds and Proportion Correct in Experiment 1.

	Believable	Unbelievable	Neutral
Log Odds	0.76 (0.16)	0.87 (0.24)	0.71 (0.16)
Proportion Correct	.68	.70	.67

*Note:* Standard errors in parenthesis. Proportions correct are marginal means calculated on the basis of the log odds. Contrast coding was used such that each log odds is calculated based on deflection from the grand mean.

**Table 3**  
Examples of the non-conflict and conflict problem types used in Experiment 2.

Problem type	Valid syllogism	Invalid syllogism
Non-conflict	Some animals are piffures No piffures are dogs Some animals are not dogs [B]	Some animals are piffures No piffures are dogs Some dogs are not animals [U]
Conflict	Some dogs are piffures No piffures are animals Some dogs are not animals [U]	Some dogs are piffures No piffures are animals Some animals are not dogs [B]

Note: In the actual Experiment all four figures and both conclusion directions were used to ensure that both sets of premises were not always identical. The valid argument was presented on the left half the time and on the right half the time. [B] = believable conclusion, [U] = unbelievable conclusion.

difference in accuracy between the unbelievable and believable conditions. This null result is more difficult to interpret than a positive result because it lends itself to several possibilities. It might be that having to think about two arguments side by side causes people to ignore beliefs altogether. This would be surprising given that pre-existing beliefs seem to be so difficult to ignore when arguments are presented in isolation. Nevertheless, the implications are interesting and we will consider in due course how forced choice judgments could alter reasoning strategies. A less dramatic possibility, one consistent with the findings of Dube et al. (2010), is that conclusion believability only affects response bias.<sup>2</sup> If this were the case, one would expect identical accuracy for believable and unbelievable trials. This possibility will be examined in the next experiment.

### 3. Experiment 2

In Experiment 1, we ensured that logic and belief were uncorrelated, eliminating the possibility that a simple belief-driven response bias could influence responding. In this experiment we reintroduced this possibility by creating two problem types in which logic and belief were perfectly correlated. For the non-conflict problems, logic and belief were perfectly positively correlated: the valid syllogism always had a believable conclusion and the invalid syllogism always had an unbelievable conclusion (top row of Table 3). For the conflict problems, logic and belief were perfectly negatively correlated: the valid syllogism always had an unbelievable conclusion and the invalid syllogism always had a believable conclusion (see bottom row of Table 3). If people are biased toward endorsing believable conclusions, the proportion of correct choices should be higher in the non-conflict condition where the

believable conclusion is always valid, relative to the conflict condition where the believable conclusion is always invalid.

#### 3.1. Method

##### 3.1.1. Participants

A total of 23 participants took part in exchange for course credit. Four participants were male and 19 were female (age 18–25,  $M = 20$ ,  $SD = 2$ ).

##### 3.1.2. Materials

Materials were constructed according to the same principles as in Experiment 1, with the exception that the non-conflict and conflict problem types were substituted for the believable and unbelievable ones. Non-conflict problems consisted of a valid–believable syllogism presented alongside an invalid–unbelievable one. Conflict problems contained a valid–unbelievable syllogism presented alongside an invalid–believable one. The neutral problems were included for consistency with Experiment 1. A unique problem list containing 48 problems (16 non-conflict, 16 conflict, 16 neutral) was constructed for each participant by randomly assigning item contents to syllogistic structures.

##### 3.1.3. Procedure, measures, and design

The procedure and measures were the same as in Experiment 1. The participant's choice of argument (left vs. right) and the accuracy of the choice (whether the valid response was chosen) was recorded on every trial. The experiment had a one-way design with problem type (non-conflict vs. conflict vs. neutral) manipulated within subjects.

#### 3.2. Results

##### 3.2.1. Preliminary analyzes

To rule out bias based on spatial position, we fit an intercept only logit GLMM with left selections as the dependent variable (dummy coded 1 = left, 0 = right). There was no evidence for spatial bias,  $LO = 0.03$ ,  $z = 0.48$ ,  $p = .63$ . Participants chose the left argument 51% of the time.

We tested whether participants responded above chance using an intercept only logit GLMM with correct responses as the dependent variable. Performance was significantly above chance,  $LO = 0.87$ ,  $z = 6.16$ ,  $p < .001$ . Mean proportion correct was 70%.

##### 3.2.2. Non-conflict vs. conflict

A summary of performance can be found in Table 4. To test for the possibility of response bias we performed a logit GLMM analysis on correct responses with problem type (non-conflict vs. conflict) as the predictor. There was no effect of problem type on correct responses,  $\chi^2(1) = 0.41$ ,  $p = .52$ .

Performance on the non-conflict and conflict problem types was compared with the neutral baseline. Logit GLMM (problem type: non-conflict vs. conflict vs. neutral) on correct responses revealed no difference between the problem types,  $\chi^2(2) = 0.41$ ,  $p = .81$ .

<sup>2</sup> It is well known in SDT that under certain conditions, evidence can change without altering judgment accuracy, resulting in performance similar to a shift in response bias. When describing the data, we refer to patterns as being consistent or inconsistent with response bias, but consider the alternative interpretation in the final discussion.

**Table 4**  
Log Odds and Proportion Correct in Experiment 2.

	Non-conflict	Conflict	Neutral
Log Odds	1.03 (0.20)	0.79 (0.27)	0.92 (0.18)
Proportion Correct	.74	.69	.71

Note. Standard errors in parenthesis. Proportion correct is calculated on the basis of the log odds.

### 3.3. Discussion

Although participants were clearly engaged in the task (accuracy was well above chance), we again found no evidence for an effect of beliefs on endorsement rates. In Experiment 1, participants did not reason more accurately when faced with unbelievable conclusions. In Experiment 2, participants displayed no preference for believable conclusions. These findings are surprising given that the motivation of the experiments was to follow up on previous ROC work reporting accuracy and response bias effects (Dube et al., 2010; Trippas et al., 2013). At first glance, they also seem to be at odds with most extant theories of belief bias; we will examine this more closely in the final discussion. The most obvious explanation for the null results is that the forced choice format led participants to ignore beliefs altogether. This may seem improbable given the known difficulty of ignoring beliefs when arguments are presented in isolation. However, support for this explanation might be found in selective processing theory (e.g., Evans et al., 2001; Klauer et al., 2000) which argues that reasoning proceeds from the conclusion to the premises. Manipulations that encourage reasoning from premises to conclusion are known to reduce belief bias effects (Morley, Evans, & Handley, 2004). Perhaps the simultaneous presentation of two adjacent syllogisms in the forced choice format encourages a focus on syllogistic structure rather than problem content. This results in a premise-to-conclusion reasoning strategy in which the premises of both arguments are compared and processed simultaneously. We examined this possibility in Experiment 3.

## 4. Experiment 3

In this experiment, the forced choice procedure was modified to induce a conclusion-to-premise reasoning strategy in the hopes of making the reasoning process more similar to the single probe method traditionally used to study belief bias. In the previous experiments, participants were presented with two complete syllogisms, each composed of two premises and a conclusion, at the onset of the trial. In Experiment 3, participants were initially presented with two conclusions in the absence of the premises. The participant then had to select one conclusion in order to reveal its premises, forcing participants to reason about each argument in turn. The non-simultaneous presentation would make it difficult to directly compare the structures of the two problems. If the absence of belief effects in the previous experiments were due to the procedural artefact of simultaneous presentation, we should now find an effect of beliefs.

All of the critical problem types from the previous experiments were included in this experiment. Our primary interest was in the effect of beliefs on the quality of reasoning. If unbelievable conclusions encourage more effective reasoning, accuracy should be higher for the unbelievable compared to the believable problems. In the absence of a difference between the unbelievable and believable conditions, a comparison of the non-conflict and conflict problems allows a test for belief-based response bias. A bias toward accepting believable conclusions would produce higher accuracy in the non-conflict compared to the conflict problems.

### 4.1. Method

#### 4.1.1. Participants

A total of 53 people (18 of which were male) took place in exchange for course credit (age: range = 18–31,  $M = 20$ ,  $SD = 3$ ).

#### 4.1.2. Materials

Materials were constructed as in Experiments 1 and 2. Participants were presented with all four critical problem types (believable, unbelievable, non-conflict, conflict) for a total of 64 problems (16 of each type). The problems were presented in a fully randomized order, with unique randomisation for each participant.

#### 4.1.3. Procedure, measures, and design

The procedure was identical to the procedure in the previous experiments with the exception that the arguments were no longer presented simultaneously. On each trial, participants were presented with two boxes containing only a conclusion. Upon clicking one of the boxes, the premises associated with that syllogism appeared, allowing the participant to reason about its validity while the premises of the other syllogism remained hidden. Clicking the other box revealed the premises associated with that syllogism and made the premises of the other syllogism disappear. Participants were therefore only able to see one problem at a time. The participants were allowed to switch between the two problems as often as they desired. The recorded response was the box that was ultimately selected. The participant's choice of argument (left vs. right) and the accuracy of the choice was recorded on every trial. Problem type (non-conflict vs. conflict; believable vs. unbelievable) was manipulated within subjects.

### 4.2. Results

#### 4.2.1. Preliminary analyses

To rule out bias based on spatial position, we fit an intercept only logit GLMM with left selections as the dependent variable (dummy coded 1 = left, 0 = right). There was no significant preference toward either side,  $LO = -0.05$ ,  $z = -1.51$ ,  $p = .13$ . The average proportion of left selections was 49%.

We tested whether the participants reasoned above chance using logit GLMM analysis on correct responses (correct = 1, incorrect = 0). Performance was above chance,

**Table 5**  
Log Odds and Proportion Correct in Experiment 3.

	Believable	Unbelievable	Non-conflict	Conflict
Log Odds	1.00 (0.12)	1.44 (0.17)	1.31 (0.14)	1.16 (0.15)
P(Correct)	.73	.81	.79	.76

Note: Standard errors in parenthesis. Proportion correct is calculated on the basis of the log odds.

$LO = 1.17$ ,  $z = 11.28$ ,  $p < .001$ . Average proportion correct was 76%.

#### 4.2.2. Believable vs. unbelievable

A summary of performance can be found in Table 5. We tested for qualitative effects of beliefs on reasoning by comparing proportion correct between the believable and unbelievable problem types using a logit GLMM with problem type (believable vs. unbelievable) as predictor. There was a significant effect of problem type on correct responses  $\chi^2(1) = 5.70$ ,  $p = .017$ .

#### 4.2.3. Non-conflict vs. conflict

A comparison of the non-conflict and conflict problem types using a logit GLMM revealed no significant effect of problem type on correct responses,  $\chi^2(1) = 0.54$ ,  $p = .46$ .

### 4.3. Discussion

Modifying the forced choice procedure so that premises were no longer presented simultaneously resulted in significantly greater accuracy for unbelievable than for believable trials, the first indication so far of an effect of beliefs. The results provide converging evidence for Trippas et al.'s (2013) finding, based on the analysis of ROCs, that belief bias involves a change in the quality of reasoning and is not solely the product of response bias. Equally interesting is what the results imply about the forced choice method. The divergent findings of Experiments 1 and 3 reveal the importance of presentation format in shaping reasoning strategies. We hypothesized that having both arguments fully available at once encourages participants to focus on argument structure, comparing them from premise to conclusion in a manner known to reduce belief bias (Morley et al., 2004). Avoiding the influence of beliefs on reasoning is notoriously difficult, and a presentation format that achieves this is noteworthy.

Our primary interest, however, lies with the results achieved in Experiment 3 using non-simultaneous presentation. Having to consider the arguments one at a time produced a belief bias effect more in keeping with that observed when arguments are judged in isolation. One could argue, of course, that the source of the effect in the two tasks is not necessarily the same. After all, the null results of Experiments 1 and 2 show that forced choice can alter the way that people reason. In the remaining experiments, we sought not only to replicate these results but to find additional convergence with belief bias effects observed using traditional methods.

## 5. Experiment 4

The results of Experiment 3 provide new evidence that beliefs have a qualitative effect on the reasoning process. However, theories of belief bias that predict such an effect have largely been developed in the context of studies using the traditional method of presenting arguments in isolation. In order to establish the relevance of our findings to the larger body of theory, it would be desirable to provide further evidence of the similarity between forced choice results and those produced with traditional methods. Trippas et al. (2013) found that not only did beliefs affect the accuracy of validity judgments but that this was mediated by individual differences. Participants of higher cognitive ability reasoned more accurately when faced with unbelievable conclusions. Those of lower cognitive ability failed to show this benefit although, consistent with a response bias, they were more accepting of believable conclusions. The pattern makes intuitive sense: enhanced reasoning strategies require more sophisticated knowledge or greater motivation while a heuristic bias requires neither. Experiment 4 aimed to replicate the results of Experiment 3 with an added test of cognitive ability. Providing new evidence for the role of individual differences in belief bias would be interesting by itself, but it would also further establish that the reasoning process involved in the forced choice task used here is similar to that involved in making judgments about isolated arguments.

### 5.1. Method

#### 5.1.1. Participants

A total of 108 undergraduate psychology students (17 male) from Plymouth University volunteered to take part in the experiment in exchange for course credit (age: range = 18–40,  $M = 20$ ,  $SD = 4$ ).

#### 5.1.2. Materials

The materials were constructed in the same way as in Experiment 3. All participants were presented with believable, unbelievable, non-conflict, and conflict problems for a total of 64 trials to test for qualitative belief effects and response bias.

Cognitive ability was measured using the AH4 group test of general intelligence (Heim, 1970). This test has been shown to be linked to deductive reasoning and belief bias (Newstead, Handley, Harley, Wright, & Farrelly, 2004; Trippas et al., 2013).

#### 5.1.3. Procedure, measures, and design

The procedure was very similar to that of the previous experiment with the exception that cognitive ability was measured prior to the reasoning task.

Following Trippas et al. (2013), we performed a median split on cognitive ability to divide our sample into a higher cognitive ability group and a lower cognitive ability group (the results did not change when cognitive ability was included as a continuous predictor, see Appendix C). For the reasoning task, the participant's choice of argument (left vs. right) and the accuracy of the choice was recorded



on every trial. Problem type (non-conflict vs. conflict; believable vs. unbelievable) was manipulated within subjects, and cognitive ability (high vs. low) was a between-subjects factor.

## 5.2. Results

### 5.2.1. Preliminary analyses

To rule out bias based on spatial position, we fit an intercept only logit GLMM with left selections as the dependent variable (dummy coded 1 = left, 0 = right). Participants had no preference for either box,  $LO = -0.04$ ,  $z = -1.57$ ,  $p = .12$ . The left box was chosen approximately 49% of the time.

We tested whether participants reasoned above chance using an intercept only logit GLMM with correct responses (1 = correct, 0 = incorrect) as predictor. Participants performed significantly above chance,  $LO = 0.92$ ,  $z = 12.39$ ,  $p < .001$ . On average, participants selected the correct response 72% of the time.

### 5.2.2. Believable vs. unbelievable

A summary of performance can be found in Table 6. Using a logit GLMM, we regressed correct responses (1 = correct, 0 = incorrect) on problem type (believable vs. unbelievable), cognitive ability (higher vs. lower), and the interaction between both. Accuracy was higher for the higher cognitive ability group,  $\chi^2(1) = 8.55$ ,  $p = .003$ . There was no main effect of believability,  $\chi^2(1) = 0.06$ ,  $p = .80$ . However, problem type and cognitive ability interacted, indicating that the effect of problem type differed for the higher and lower ability groups,  $\chi^2(1) = 6.42$ ,  $p = .011$ . Follow up tests indicated that the higher ability participants reasoned better for the unbelievable problems compared to the believable problems,  $\chi^2(1) = 5.66$ ,  $p = .017$ . The lower ability participants showed a trend in the opposite direction,  $\chi^2(1) = 3.26$ ,  $p = .07$ .

### 5.2.3. Non-conflict vs. conflict

Using logit GLMM, we regressed correct responses (1 = correct, 0 = incorrect) on problem type (non-conflict vs. conflict), cognitive ability (higher vs. lower), and the interaction between both. A main effect of problem type demonstrated that accuracy was higher for the non-conflict compared to the conflict problems,  $\chi^2(1) = 11.42$ ,  $p < .001$ . A main effect of cognitive ability demonstrated that accuracy was higher for the higher ability group,  $\chi^2(1) = 13.59$ ,  $p < .001$ . Conflict and ability did not interact,  $\chi^2(1) = 1.44$ ,  $p = .23$ . Interpreting the effect of conflict is not straightforward when beliefs are shown

to affect accuracy. For the lower ability group, which did not show an accuracy effect, these results reveal that beliefs still exert an influence, albeit in a manner consistent with pure response bias.

## 5.3. Discussion

We replicated the finding from Experiment 3 of enhanced reasoning for unbelievable conclusions. This facilitation was observed only for the higher cognitive ability group. For the lower ability group, conclusion believability did not enhance reasoning accuracy, although performance during non-conflict and conflict trials was consistent with a bias to endorse believable conclusions. The full pattern of results correspond with the findings of Trippas et al. (2013), strengthening the argument that the non-simultaneous forced choice judgment engages the reasoning process in much the same way as the traditional single probe judgment. More important to our understanding of belief bias, the results offer independent support for the claim that beliefs alter the quality of reasoning in a manner mediated by individual differences.

## 6. Experiment 5

Trippas et al. (2013) speculated that the tendency to engage in belief-driven enhanced reasoning strategies might be as much a matter of motivation as of ability. Analytic cognitive style is another aspect of personality thought to play an important role in reasoning (Toplak, West, & Stanovich, 2011, 2014). Cognitive style encompasses not only the capacity to engage in analytic processing but also the motivation to do so. In Experiment 5, we extended our investigation of individual differences to the effect of cognitive style on belief bias.

We also scrutinized the effect of beliefs on the lower cognitive ability group which was apparent only in the non-conflict and conflict trials. The trend in Experiment 3, although in the same direction, was small and not significant. Because this may have been due to aggregating across ability groups, it would be good to replicate the clear pattern that emerged in Experiment 4 when individual differences were taken into account. An additional manipulation was introduced to test the plausibility of characterizing the pattern as a response bias. Response heuristics are often strategic in nature (Gigerenzer & Brighton, 2009). Although the preference for accepting believable conclusions may be deeply rooted in experience, it may also be attuned to local contingencies. A problem with mixing non-conflict and conflict problems together

**Table 6**

Log Odds and Proportion Correct as a function of cognitive ability of Experiment 4.

		Believable	Unbelievable	Non-conflict	Conflict
Higher CA	Log Odds	1.17 (0.12)	1.61 (0.22)	1.37 (0.13)	1.08 (0.20)
	P(Correct)	.76	.82	.80	.75
Lower CA	Log Odds	0.78 (0.09)	0.56 (0.11)	0.95 (0.10)	0.29 (0.13)
	P(Correct)	.69	.64	.72	.57

Note: Standard errors in parenthesis. Proportion correct is calculated on the basis of the log odds.

with believable and unbelievable problems is that with the latter, which account for half of the trials in the experiment, responding based on conclusion believability alone is not possible. This may cause participants to discard their otherwise habitual use of a simple, belief-based heuristic. To examine this possibility, we blocked the conditions. Some participants were exposed to only non-conflict and conflict trials in the first half of the list and only believable and unbelievable trials in the second half. Other participants were exposed to the conditions in the reverse order. If exposure to believable and unbelievable trials is responsible for suppressing a bias to accept believable conclusions, the accuracy difference between non-conflict and conflict problems should be most pronounced when participants are initially exposed to only these problems.

## 6.1. Method

### 6.1.1. Participants

A total of 71 undergraduate psychology students (8 male) from Plymouth University volunteered to take part in the experiment in exchange for course credit (age: range = 18–54,  $M = 23$ ,  $SD = 7$ ).

### 6.1.2. Materials

The materials were constructed as in Experiments 3 and 4. All participants were presented with believable, unbelievable, non-conflict, and conflict problems for a total of 64 trials to test for enhanced reasoning and consistency with response bias.

Analytic cognitive style (ACS) was tested using the CRT (Frederick, 2005), considered to be a short and efficient performance measure of analytic cognitive style (e.g., Pennycook, Cheyne, Seli, Koehler, & Fugelsang, 2012; Pennycook, Cheyne, Koehler, & Fugelsang, 2013). The test involves answering three questions which cue an intuitively compelling (but incorrect) response. For instance: “If it takes 5 machines 5 min to make 5 widgets, how long does it take 100 machines to make 100 widgets?” The intuitive response in this case is to use a matching heuristic and to respond “100.” However, participants with adequate levels of analytic cognitive style will realize that this is not the case, given that it takes one machine five minutes to make one widget. The ability to inhibit the intuitive responses in the CRT has been linked to rational thinking ability and actively open-minded thinking, two alternative measures of analytic cognitive style (Campetilli & Gerrans, 2013; Toplak et al., 2011, 2014).

### 6.1.3. Procedure, measures, and design

The procedure was identical to that of the previous experiments with the exception that upon completing the experiment, the participants were asked to complete the CRT. Participants were assigned to the lower analytic cognitive style group if they did not correctly solve any of the three CRT questions ( $n = 42$ ). For the reasoning task, the participant’s choice of argument (left vs. right) and the accuracy of the choice was recorded on every trial.

The design differed from the previous experiments in that the problem types were blocked. For half of the participants, the first 32 problems consisted of only believable

and unbelievable trials and the final 32 problems consisted of only non-conflict and conflict trials. For the remaining participants, the reverse was true. Participants were not informed about the blocked nature of this design and there was no pause in between the lists. The intent of the design modification was to control for the influence of the believable vs. unbelievable trials on heuristic strategies adopted for non-conflict vs. conflict trials. Problem type (non-conflict vs. conflict; believable vs. unbelievable) was manipulated within subjects, and analytic cognitive style (higher vs. lower) was a between-subjects factor.

## 6.2. Results

### 6.2.1. Preliminary analyzes

To rule out bias based on spatial position, we fit an intercept only logit GLMM with left selections as the dependent variable (dummy coded 1 = left, 0 = right). There was no spatial bias,  $LO = -0.05$ ,  $z = -1.56$ ,  $p = .12$ . Approximately 49% of the times the box on the left was selected.

To test whether the participants reasoned above chance we fit an intercept only logit GLMM to correct responses (correct = 1, incorrect = 0). Participants performed significantly above chance,  $LO = 0.82$ ,  $z = 8.97$ ,  $p < .001$ . The probability of giving the correct response was 69% on average.

### 6.2.2. Believable vs. unbelievable

A summary of performance can be found in Table 7. We analyzed correct responses (correct = 1, incorrect = 0) using a GLMM logit regression model with a full factorial 2 (problem type: believable vs. unbelievable)  $\times$  2 (analytic cognitive style: higher vs. lower)  $\times$  2 (block order: non-conflict vs. conflict trials presented first vs. presented last) design. Analytic cognitive style and problem type interacted,  $\chi^2(1) = 11.73$ ,  $p < .001$ . Follow up tests demonstrated that beliefs increased reasoning accuracy for the higher ACS group,  $\chi^2(1) = 6.54$ ,  $p = .011$ . There was no effect of believability for the lower ACS group,  $\chi^2(1) = 2.14$ ,  $p = .14$ . No other effects were significant, all  $ps > .12$ .

### 6.2.3. Non-conflict vs. conflict

A summary of performance can be found in Table 8. Correct responses were analyzed using a GLMM logit regression with a full factorial 2 (problem type: non-conflict vs. conflict)  $\times$  2 (analytic cognitive style: higher vs. lower)  $\times$  2 (block order: non-conflict and conflict trials presented first vs. presented last) design. The main effect

**Table 7**

Log Odds and Proportion Correct as a function of analytic cognitive style for unbelievable and believable problems in Experiment 5.

		Believable	Unbelievable
Higher ACS	Log Odds	0.93 (0.16)	1.44 (0.22)
	Proportion Correct	.71	.81
Lower ACS	Log Odds	0.82 (0.13)	0.61 (0.17)
	Proportion Correct	.70	.65

Note: Standard errors in parenthesis. Proportion correct is calculated on the basis of the log odds.

**Table 8**

Log Odds and Proportion Correct as a function of analytic cognitive style and presentation order of non-conflict and conflict problems in Experiment 5.

		Presented first		Presented last	
		Non-conflict	Conflict	Non-conflict	Conflict
Higher ACS	Log Odds	1.64 (0.36)	1.32 (0.34)	1.02 (0.22)	0.86 (0.35)
	<i>P</i> (correct)	.84	.79	.73	.70
Lower ACS	Log Odds	2.09 (0.36)	−0.29 (0.40)	0.63 (0.13)	0.20 (0.19)
	<i>P</i> (correct)	.89	.43	.65	.55

Note: Standard errors in parenthesis. Proportion correct is calculated on the basis of the log odds.

of problem type indicated that accuracy was higher for the non-conflict compared to the conflict problems,  $\chi^2(1) = 10.55, p = .001$ . The main effect of analytic cognitive style indicated that accuracy was higher for the higher ACS group,  $\chi^2(1) = 4.77, p = .029$ . The main effect of block order indicated that accuracy was higher overall when the non-conflict and conflict trials were presented first,  $\chi^2(1) = 5.94, p = .015$ . These main effects were qualified by several interactions. Problem type and analytic cognitive style interacted,  $\chi^2(1) = 6.68, p = .001$ , suggesting that the effect of problem type was larger for the lower ACS group. Problem type and block order also interacted,  $\chi^2(1) = 5.46, p = .019$ , suggesting that the effect of problem type was more pronounced if the non-conflict and conflict trials were presented first. The three-way interaction between problem type, analytic cognitive style, and block order was also significant,  $\chi^2(1) = 4.40, p = .036$ . This led us to perform follow up analyses for the higher and lower cognitive style groups separately.

For the higher ACS group, the 2 (problem type)  $\times$  2 (block order) GLMM logit regression did not produce any effects, all *ps* > .27. For the lower ACS group, the same analysis resulted in main effects of problem type,  $\chi^2(1) = 11.92, p < .001$ , and block order,  $\chi^2(1) = 6.64, p = .010$ . Finally, problem type and block interacted,  $\chi^2(1) = 8.63, p = .003$ , showing that the effect of problem type was larger when the non-conflict and conflict trials were presented first.

### 6.3. Discussion

Replicating the findings of the previous two experiments, participants reasoned more accurately when faced with unbelievable conclusions. Analytic cognitive style mediated this effect. Similar to the pattern observed with cognitive ability in Experiment 4, only the higher ACS group showed enhanced reasoning on unbelievable trials. The effect may be due in part to higher ACS individuals possessing greater cognitive ability. However, as De Neys, Rossi, and Houdé (2013) observed, many people still choose an intuitively compelling response even when they suspect it may be incorrect. Higher ACS participants may possess not only the knowledge to reason effectively but also the motivation or impulse control to ignore easier but less accurate strategies.

In Experiment 4, the lower cognitive ability group showed no difference in the quality of their reasoning about believable and unbelievable problems. However, beliefs did influence their performance during non-conflict and conflict trials in a manner consistent with a bias to

accept believable conclusions. In Experiment 5, the lower ACS group showed a similar pattern of performance during non-conflict and conflict trials. We suspected that the use of simple, belief-based heuristics, while habitual, may also be sensitive to local contingencies. Non-conflict and conflict problems make responding based on conclusion believability possible and even reasonable. Believable and unbelievable problems, on the other hand, remove the heuristic as an option and in doing so may discourage its use. We hypothesized that participants who were initially given only non-conflict and conflict trials would be more likely to use heuristic responding than those who began the session with unbelievable and believable trials. Therefore, the finding that differential responding to non-conflict and conflict problems was more pronounced when these trials were presented first rather than last in the session suggests to us that response bias is driving this effect. Interestingly, the improved reasoning about unbelievable problems of the higher ACS group did not depend on presentation order. This suggests that the critical effects observed in the non-conflict vs. conflict trials and unbelievable vs. believable trials are tied to different mechanisms.

## 7. General discussion

A number of competing theories have arisen over the years to explain belief bias, the influence of pre-existing beliefs on judgments of logical validity. The predominant view has been that belief bias has two components. Response bias, the more basic component, is the tendency to accept believable conclusions as valid, regardless of their actual validity. Response bias describes a simple heuristic that influences endorsement rates without affecting the accuracy of validity judgments. Beliefs may also influence the quality of reasoning. Although theories differ on the specific mechanism, the major theories suggest that people tend to reason more diligently, and thus more accurately, when confronted with unbelievable conclusions. Evidence for a qualitative effect of beliefs on reasoning has traditionally focused on the interaction between conclusion believability and validity that is often observed in argument endorsement rates. However, recent work taking a formal modeling approach has shown that interpreting such interactions requires assumptions whose empirical and theoretical justification is open to question (Dube et al., 2010; Heit & Rotello, 2014; Klauer et al., 2000). Applying a SDT model to their own syllogistic reasoning data, Dube et al. found no evidence for enhanced

reasoning. Furthermore, the SDT model demonstrated that belief  $\times$  logic interactions could arise solely as a consequence of changes in response bias.

### 7.1. Theoretical implications

The work critical of traditional methods of analyzing belief bias, in particular Dube et al.'s (2010) argument that pure response bias account offers a viable alternative to more complex theories, suggests a need to reconsider the evidence for a dual component view of belief bias. Trippas et al. (2013), using ROC analysis based on the SDT model, found that beliefs did affect the accuracy, and therefore the quality, of syllogistic reasoning under certain conditions. When the task discouraged extended analytic processing by using simple problems or time pressure, belief bias could be accounted for purely in terms of a simple response bias. The qualitative effect of beliefs emerged with complex problems, self-paced responding, and most interestingly, only in the subset of participants scoring higher on measures of cognitive ability. In a sense, these findings can accommodate both the dual component and pure response bias accounts if it is assumed that each holds under conditions. There is reason, however, to view this conclusion as preliminary. Although the analysis of confidence rating-based ROCs in reasoning has strong advocates (e.g., Pazzaglia, Dube, & Rotello, 2013), there remains debate about its use (Klauer & Kellen, 2011; Dube et al., 2011). Trippas et al. (2013) attributed the conflicting results of their study and that of Dube et al. (2010) to individual differences causing variation across samples. An explanation based on individual differences requires replication and independent verification.

Two-alternative forced choice is a task with properties that are well understood within the framework of SDT. This makes it a natural vehicle for seeking converging evidence with prior work using ROC methods. That said, the ability to control response bias in forced choice makes it a useful tool to test for effects of beliefs on reasoning regardless of theoretical perspective. In Experiments 3–5, we confirmed that beliefs do influence the quality of reasoning. We also showed that enhanced reasoning for unbelievable conclusions depends on individual differences such as cognitive ability and analytic cognitive style. In sum, our findings replicate and extend the findings of Trippas et al. (2013) using this novel methodology.

If response bias alone is not sufficient to account for belief bias, which of the traditional theories can best account for our findings? Mental models theory (MMT; Oakhill et al., 1989) explicitly proposes a premises-to-conclusion reasoning strategy in which a tentative conclusion is generated from the premises, with an additional search for counterexamples if that initial conclusion is unbelievable. Under this account, we would have expected an accuracy advantage for unbelievable problems in Experiment 1. Selective processing theory (SPT; Evans et al., 2001; Klauer et al., 2000) fares slightly better in explaining the discrepancy between Experiments 1 and 3. According to SPT, belief bias emerges from a conclusion-to-premises reasoning process. Participants first

evaluate the believability of the premises, after which they conduct either a positive test if the conclusion is believable or a negative test if it is unbelievable. For complex indeterminately invalid syllogisms of the type used in our experiments, multiple representations of the premises exist (some of which are compatible with the premises, others incompatible). As such, the negative test cued by the unbelievable conclusion will lead to increased rejection of invalid arguments, resulting in higher judgment accuracy. Conversely, the positive test cued by believable conclusions will lead to the increased acceptance of invalid arguments, resulting in lower judgment accuracy. For valid complex syllogisms, only compatible models exist, and consequently alternate belief-based testing strategies do not result in any differences in judgment accuracy. SPT does not explicitly take individual differences into account and so has little to say about their role in the findings from Experiments 4 and 5. There has been one previous attempt to extend SPT to individual differences. Stuppel, Ball, Evans, and Kamal-Smith (2011) used latency analysis to show that distinct subgroups existed among reasoners in a belief bias task. One subgroup appeared to operate in a manner wholly consistent with response bias, while other groups showed enhanced reasoning or perfectly normative reasoning.

Our findings are also compatible with dual process theory, which suggests that enhanced reasoning is driven by effortful, reflective Type 2 (T2) processing. A key determinant of T2 processing is its link with the recruitment of working memory. Higher ability participants presumably have a larger working memory capacity, which is consistent with the idea that those of higher cognitive ability engage more effortful reasoning under certain conditions. This focus on individual differences resonates with suggestions made by other proponents of dual process theory (e.g., De Neys & Bonnefon, 2013). The two components of belief bias have often been cited as one of the key pieces of evidence for the dual process view (Evans & Stanovich, 2013). According to DPT, the response bias component of belief bias is driven by Type 1 (T1) processing and the accuracy component is a marker for T2 processing. With respect to the current findings, DPT would predict that under conditions that discourage effortful processing, such as working memory load or response deadlines, the effect of beliefs on accuracy should disappear even in those with higher cognitive ability or ACS. The effect of beliefs on heuristic bias, in contrast, should be unaffected or even facilitated.

The claim that beliefs influence the quality of reasoning lies at the heart of the various theories described above. In using the forced-choice procedure, our aim was to provide clear evidence of this using methodology that can control for changes in performance due to response bias. Therefore, performance in the unbelievable vs. believable trials make up the critical results of the study. However, it is a prevalent, albeit usually informal, assumption that people are often biased toward accepting believable arguments regardless of their validity. As a secondary aspect of our investigation, we used performance in the non-conflict vs. conflict trials to test

for the presence of response bias when beliefs failed to affect judgment accuracy. We should be clear, however, that although response bias would produce an accuracy advantage for non-conflict problems, there is another possible explanation for this result. Imagine that believable conclusions, regardless of validity, tended to possess more “evidence strength” than unbelievable conclusions. Under certain conditions, this might not change the ability to judge validity (i.e., accuracy would be the same for unbelievable and believable trials). However, because believable conclusions possess more evidence strength, participants would be more likely to choose them in non-conflict and conflict problems. Therefore, the effects observed in non-conflict vs. conflict trials in the later experiments could be characterized in this way rather than as response bias. Although this possibility is worth investigating, the import of the non-conflict vs. conflict results does not really depend on certainty about the underlying mechanism. Rather, the results are interesting because they reveal that lower cognitive ability and analytic cognitive style participants are being influenced by beliefs, but in a seemingly different manner than higher ability and analytic cognitive style participants. This is not to say that differentiating response bias from other mechanisms is impossible. One can, for example, use manipulations that plausibly target response bias. We suggest that the vulnerability of the non-conflict vs. conflict effect to presentation order, observed in Experiment 5, is what would be expected from response bias.

## 7.2. Additional implications

Although Experiments 3–5 contribute the central findings of the study, the null findings of Experiments 1 and 2 yielded an unexpected but potentially significant insight into the nature of belief bias. Participants performed adequately in the reasoning task, but there was no evidence that beliefs affected the quality of reasoning, nor was there any evidence for a belief-based response bias. We considered the possibility that the force choice method itself removes the influence of beliefs on reasoning. However, belief effects emerged in later experiments following a slight modification of the procedure. In Experiments 1 and 2, two complete syllogisms were presented side by side. This allowed participants to directly compare the structure of the two arguments. Comparison might naturally proceed from top to bottom, in other words, from the premises to the conclusion, given that the only differences between the arguments are present in the premises. [Morley et al. \(2004\)](#) noted that premise-to-conclusion reasoning tends to be immune to belief bias. In Experiments 3–5, the procedure was modified so that the premises of the two problems were no longer simultaneously visible. Instead, both conclusions were presented and participants had to select each to view its accompanying premises. This resulted in belief bias effects in Experiments 3–5 analogous to those observed with the traditional, single-probe procedure.

Why would the simultaneous-presentation procedure have such a different effect on reasoning? Work by

[Pachur and Olsson \(2012\)](#) on knowledge acquisition may be informative. In one condition, learning consisted of viewing each object in isolation and making a judgment about its characteristics (e.g., poison level). In another condition, learning involved viewing objects in pairs and making a judgment about their relative characteristics (e.g., which was more poisonous). Participants who experienced single-probe learning later tended to judge characteristics using an exemplar-based strategy, comparing new objects to previously encountered objects. Those who experienced forced-choice learning, on the other hand, later tended toward more abstract, rule-based judgments. Applied to syllogistic reasoning, it may be that when problems are presented in isolation, people are prone to judge validity based on prior experience, analogous to an exemplar strategy. Prior experience is presumably the source of belief bias. When problems are compared side by side, people may instead tend toward rule-based judgments, i.e., formal logic.

In reasoning and decision-making, the influence of irrelevant beliefs is widespread and difficult to avoid. Instructing people to ignore irrelevant information and reason deductively is rarely effective ([Evans, Newstead, Allen, & Pollard, 1994](#); [Heit & Rotello, 2014](#)). If the goal is to measure reasoning ability, our findings suggest that controlling presentation format may be a more effective way to limit extraneous influences. As a didactic tool, forced choice presentation may prove to be more effective than the traditional method of presenting problems or examples in isolation. In academic settings, the focus is typically on learning to apply formal rules. Using side-by-side presentation in which irrelevant factors are equated between the examples draws attention to the relevant factors and also draws attention to broader, structural regularities, both of which likely contribute to findings like those of [Pachur and Olsson \(2012\)](#) that forced choice learning promotes the development of rules-based strategies.

## Appendix A

See [Tables A.1 and A.2](#).

**Table A.1**  
Item contents used for reasoning materials.

Category	Members			
amphibians	frogs	salamanders	toads	newts
birds	parrots	sparrows	ducks	robins
boats	kayaks	canoes	yachts	speedboats
cars	BMW's	Volvos	Vauxhalls	Fiats
criminals	robbers	murderers	embezzlers	terrorists
furniture	desks	sofas	cupboards	bookcases
dogs	Spaniels	Labradors	Terriers	Dalmatians
drinks	beers	sodas	wines	whiskeys
fish	trout	salmons	cods	haddocks
fruits	prunes	peaches	apples	bananas
insects	bees	beetles	ants	flies
reptiles	lizards	iguanas	snakes	crocodiles
tools	hammers	saws	spanners	shovels
trees	oaks	willows	pinos	maples
vegetables	carrots	cabbages	parsnips	radishes
weapons	cannons	swords	guns	spears

**Table A.2**

Pseudowords used as linking terms in the experiments.

redes	fosks	pives	pields	decottions	sothods	renes	bunges
wasses	geets	swants	cronxes	firters	nickhomes	revoules	pinds
foins	chindles	soats	sonds	pumes	papes	trops	envenches
lebs	brops	stoges	crots	punties	stamuses	vennars	cortemns
weens	quinces	loaxes	stoals	curges	gruts	cosuors	wighes
punds	jubs	parfs	fises	hoons	tutches	brimbers	punes
cofts	spashes	fimps	brams	heets	piffures	burtes	queels
flamps	dathses	darms	vosts	trinnels	goples	boodings	veemers

**Table B.1**

Model specifications for analyses in Experiments 1, 2 and 3.

Analysis	Model
(1) Spatial bias	left~1 + (1 subj)
(2) Above chance	corr~1 + (1 subj)
(3) Believable vs. unbelievable	corr~1 + unbel + (1 + unbel subj)
(4) Non-conflict vs. conflict	corr~1 + noconf + (1 + noconf subj)
(5) Full	corr~1 + prob_type + (1 + prob_type subj)

Note: 1 = intercept, left (1 = left, 0 = right), corr (1 = correct, 0 = incorrect), subj = participant id, unbel (1 = unbelievable problem type, 0 = believable problem type), noconf (1 = non-conflict problem type, 0 = conflict problem type), prob\_type = factor (believable vs. unbelievable vs. neutral for Experiment 1; non-conflict vs. conflict vs. neutral for Experiment 2; not specified in Experiment 3). Effects between parentheses are random effects to account for the repeated measures nature of the design. The LRTs for (3), (4) and (5) were obtained by comparing with a model in which the fixed effect of interest was removed, but the random slope of that effect was retained.

**Table B.2**

Model specifications for analyses in Experiment 4.

Analysis	Model
(1) Spatial bias	left~1 + (1 subj)
(2) Above chance	corr~1 + (1 subj)
(3) Believable vs. unbelievable	corr~1 + unbel CA + (1 + unbel subj)
(4) (3) for higher CA	corr[CA==1]~1 + unbel[CA==1] + (1 + unbel[CA==1] subj[CA==1])
(5) (3) for lower CA	corr[CA==0]~1 + unbel[CA==0] + (1 + unbel[CA==0] subj[CA==0])
(6) Non-conflict vs. conflict	corr~1 + noconf + (1 + noconf subj)

Note: 1 = intercept, left: (1 = left, 0 = right), corr: (1 = correct, 0 = incorrect), subj: = participant id, unbel: (1 = unbelievable, 0 = believable), noconf: (1 = non-conflict, 0 = conflict), CA: (1 = higher cognitive ability, 0 = lower cognitive ability), an asterisk (\*) between two variables indicates that all main effects and interactions between those variables are included in the model, e.g.: unbel CA = main effect unbel, main effect CA, and unbel: CA interaction.

**Table B.3**

Model specifications for analyses in Experiment 5.

Analysis	Model
(1) Spatial bias	left~1 + (1 subj)
(2) Above chance	corr~1 + (1 subj)
(3) Believable vs. unbelievable	corr~1 + unbel ACS block + (1 + unbel subj)
(4) (3) for higher ACS	corr[ACS==1]~1 + unbel[ACS==1] block[ACS==1] + r_e1
(5) (3) for lower ACS	corr[ACS==0]~1 + unbel[ACS==0] block[ACS==0] + r_e1
(6) Non-conflict vs. conflict	corr~1 + noconf ACS block + (1 + noconf subj)
(7) (6) for higher ACS	corr[ACS==1]~1 + noconf[ACS==1] block[ACS==1] + r_e2
(8) (6) for lower ACS	corr[ACS==0]~1 + noconf[ACS==0] block[ACS==0] + r_e2
(9) (6) for lower ACS, presented first	corr[ACS==0 & block==1]~noconf[ACS==0 & block==1] + r_e2
(10) (6) for lower ACS, presented last	corr[ACS==0 & block==0]~noconf[ACS==0 & block==0] + r_e2

Note: 1 = intercept, left: (1 = left, 0 = right), corr: (1 = correct, 0 = incorrect), subj = participant id, unbel: (1 = unbelievable, 0 = believable), noconf: (1 = non-conflict, 0 = conflict), ACS: (1 = higher analytic cognitive style, 0 = lower analytic cognitive style), r\_e1 = (1 + unbel|subj), r\_e2 = (1 + noconf|subj), block (1 = non-conflict & conflict items presented first, 0 = non-conflict & conflict block presented last). LRTs for main effects and interactions were computed by comparing with a model in which the effect under scrutiny was removed.

## Appendix B

In this Appendix we provide the model specifications for all reported analyses per experiment. The models are described using the syntax as interpreted by the *glmer* function of the lme4 package in R (Bates et al., 2013). All models are estimated using maximum likelihood with a logit link function with binomially distributed residuals (see Tables B.1–B.3). LRTs for main effects and interactions were computed by comparing with a model in which the effect under scrutiny was removed. For instance, to test for the interaction between the qualitative effect of beliefs on reasoning and cognitive ability, (3) was compared with the following model (3b): corr~1 + unbel + CA + (1 + unbel|subj). To test for the main effect of enhanced reasoning, model (3b) was compared with an additional model (3c): 1 + CA + (1 + unbel|subj), and so forth.

**Table C.1**

Effects of the enhanced reasoning analysis with cognitive ability as a continuous predictor in Experiment 4.

Effect	<i>LO</i>	<i>z</i>	<i>p</i>
Intercept	0.27	0.64	.52
Cognitive ability	0.008	1.70	.09
Problem type	−1.89	−3.69	<.001
Interaction	0.002	3.79	<.001

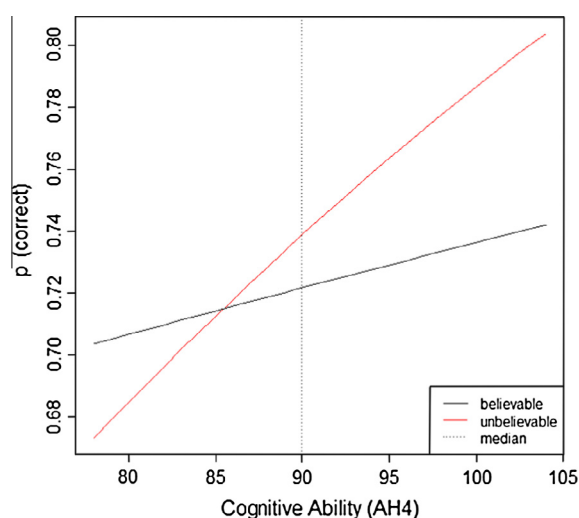
Note: Problem type: (1 = unbelievable, 0 = believable).

## Appendix C

In Experiment 4 we investigated the effect of cognitive ability on the belief-accuracy effect by comparing higher and lower cognitive ability groups using median splits. We took this approach to allow for a direct comparison with the effect of analytic cognitive style in Experiment 5, as well as with our previous work (Trippas et al., 2013; 2014). The GLMM approach allows us to inspect the effect of cognitive ability as a continuous predictor. We report this analysis below to demonstrate that this approach works equally well and leads to identical conclusions. In this analysis, treatment coding was used instead of effect coding.

We regressed correct responses (1 = correct, 0 = incorrect) on problem type (believable vs. unbelievable), cognitive ability (higher values indicate higher cognitive ability scores, range: 32–124, inter-quartile range: 78–104, median = 90), and the interaction between both. Effects are presented in Table C.1.

Based on these effects we plotted the estimated probability of getting the correct response for the believable and the unbelievable conditions as a function of cognitive ability (see Fig. C.1).



**Fig. C.1.** Performance on unbelievable and believable trials as a function of cognitive ability for Experiment 4. The vertical dotted line denotes median cognitive ability. Higher cognitive ability is associated with enhanced accuracy for unbelievable problems.

## References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D. M. (2010). *lme4: Mixed-Effects Modeling with R*. New York: Springer. <<http://lme4.r-forge.r-project.org/book/>>.
- Bates, D. M., Maechler, M., & Bolker, B. (2013). *lme4: Linear mixed-effects models using Eigen and Eigen++*. R package version 0.999999-2. <<http://CRAN.R-project.org/package=lme4>>.
- Campitelli, G., & Gerrans, P. (2013). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*. <http://dx.doi.org/10.3758/s13421-013-0367-9>.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7, 28–38. <http://dx.doi.org/10.1177/1745691611429354>.
- De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences*, 17, 172–178.
- De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, 20, 269–273.
- Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review*, 117(3), 831–863. <http://dx.doi.org/10.1037/a0019634>.
- Dube, C., Rotello, C. M., & Heit, E. (2011). The belief bias effect is aptly named: A reply to Klauer and Kellen (2011). *Psychological Review*, 118, 155–163.
- Evans, J., & 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. (2007). On the resolution of conflict in dual-process theories of reasoning. *Thinking & Reasoning*, 13, 321–329.
- 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. <http://dx.doi.org/10.1177/1745691612460685>.
- Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11, 295–306.
- Evans, J. St. B. T., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 54(A), 935–958.
- Evans, J. St. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, 6, 263–285.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1, 107–143. <http://dx.doi.org/10.1111/j.1756-8765.2008.01006.x>.
- Heim, A. W. (1970). *Group test of general intelligence*. UK: NFER-Nelson, Windsor.
- Heit, E., & Rotello, C. M. (2014). Traditional difference-score analyses of reasoning are flawed. *Cognition*, 131, 75–91.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633.
- Klauer, K. C., & Kellen, D. (2011). Assessing the belief bias effect with ROCs: Reply to Dube, Rotello, and Heit (2010). *Psychological Review*, 118, 155–164. <http://dx.doi.org/10.1037/a0020698>.
- Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, 107, 852–884. <http://dx.doi.org/10.1037/0033-295X.107.4.852>.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Morley, N. J., Evans, J. St. B. T., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, 57(4), 666–692. <http://dx.doi.org/10.1080/02724980343000440>.
- Morsanyi, K., & Handley, S. J. (2012). Logic feels so good – I like it! Evidence for intuitive detection of logic in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 38, 596–616.
- Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, 57A, 33–60.

- Newstead, S. E., Pollard, P., Evans, J. S., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*(3), 257–284.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, *31*, 117–140.
- Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology*, *65*, 207–240.
- Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*, 1173–1203.
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2013). Belief bias during reasoning among religious believers and skeptics. *Psychonomic Bulletin & Review*. <http://dx.doi.org/10.3758/s13423-013-0394-3>.
- Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), 335–346. <http://dx.doi.org/10.1016/j.cognition.2012.03.003>.
- Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *53*(4), 1202–1223.
- R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org/>>.
- Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, *34*(3), 619–632.
- Stupple, E. J. N., Ball, L. J., Evans, J. St. B. T., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology*, *23*(8), 931–941.
- Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review*, *10*, 184–189.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275–1289.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*, 147–168.
- Trippas, D., Handley, S. J., & Verde, M. F. (2013). The SDT model of belief bias: Complexity, time, and cognitive ability mediate the effects of believability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <http://dx.doi.org/10.1037/a003239>.
- Trippas, D., Handley, S. J., & Verde, M. F. (2014). Fluency and belief bias in deductive reasoning: New indices for old effects. *Frontiers in Psychology*, *5*(631), 1–7. <http://dx.doi.org/10.3389/fpsyg.2014.00631>.
- Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology*, *102*.