The SDT Model of Belief Bias: Complexity, Time, and Cognitive Ability Mediate the Effects of Believability

Dries Trippas, Simon J. Handley, and Michael F. Verde
Plymouth University

When people evaluate conclusions, they are often influenced by prior beliefs. Prevalent theories claim that belief bias affects the quality of syllogistic reasoning. However, recent work by Dube, Rotello, and Heit (2010) has suggested that belief bias may be a simple response bias. In Experiment 1, receiver operating characteristic analysis revealed that believability affected accuracy for complex but not for simple syllogisms. In Experiment 2, the effect of believability on accuracy disappeared when judgments were made under time pressure and with participants low in cognitive capacity. The observed effects on reasoning accuracy indicate that beliefs influence more than response bias when conditions are conducive to the use of certain reasoning strategies. The findings also underscore the need to consider individual differences in reasoning.

**Keywords:** reasoning, signal detection theory, belief bias, syllogisms

In everyday deductive reasoning, people often draw unwarranted conclusions when influenced by their beliefs. For example, Knauff, Budeck, Wolf, and Hamburger (2010) found that stock brokers responded less logically than meteorologists when arguments contained irrelevant content related to the stock market. The tendency for prior knowledge and beliefs to overshadow logical reasoning is known as belief bias (Evans, Barston, & Pollard, 1983). Belief bias has been studied most extensively in the domain of syllogistic reasoning (Klauer, Musch, & Naumer, 2000; Newstead, Pollard, Evans, & Allen, 1992). Consider the following syllogism:

Premise 1: No birds are balloons.

Premise 2: Some balloons are sparrows.

Conclusion: Some birds are not sparrows.

The conclusion of a syllogism is logically valid if it necessarily follows from the premises. The conclusion above is invalid because the premises do not exclude the possibility that all birds are sparrows. Nevertheless, many people accept the conclusion as valid because it seems believable (we can all think of a bird that is not a sparrow). Most current theories argue that conclusion believability affects the way that people reason. This view was recently challenged by Dube, Rotello, and Heit (2010), who proposed, based on a signal detection theory (SDT) analysis, that belief bias may be nothing more than response bias. The present study, which also takes a SDT approach, shows that belief bias is more complex than suggested by either the response bias hypothesis or other extant theories.

**Reasoning-Based Accounts**

Several theories argue that believability impacts the reasoning process by motivating people to reason more thoroughly for unbelievable conclusions. This claim is consistent with the finding that validity and conclusion believability typically interact: Endorsement rates are higher for believable than for unbelievable conclusions, but the difference is greatest for invalid conclusions. Evans et al. (1983) offered two possible explanations for this interaction. According to selective scrutiny theory, people are less likely to use formal reasoning when faced with an argument that is believable. They are more likely to accept a believable conclusion without bothering to evaluate its logical basis. For valid arguments, this belief-based heuristic and formal logic will lead to the same conclusion, thus leading to similar acceptance rates. For invalid arguments, on the other hand, many believable items will be accepted that would otherwise be rejected on the basis of logic. According to the theory of misinterpreted necessity, people are confused by some types of invalid arguments. Specifically, they are unsure how to treat indeterminately invalid arguments in which a conclusion is consistent with the premises yet does not necessarily follow from them. When faced with such an argument, people may choose to accept or reject based on the believability of the conclusion. Because this confusion arises only with invalid arguments, misinterpreted necessity does not predict an effect of believability on valid arguments. To summarize, selective scrutiny and misinterpreted necessity suggest that people are less likely to use formal reasoning.
when faced with believable conclusions. In both cases, the result is an increased tendency to accept invalid believable conclusions.

The predominant view in belief bias research is that people reason on the basis of mental models. Mental models theory (MMT; Oakhill, Johnson-Laird, & Garnham, 1989) holds that people evaluate syllogisms by constructing mental models of the premises. If the conclusion is inconsistent with the model, it is rejected. If the conclusion is consistent with the model, the believability of the conclusion is considered. A believable conclusion is immediately accepted. An unbelievable conclusion, on the other hand, encourages a search for alternative models of the premises. An unbelievable conclusion is only accepted if it remains consistent with all alternative models of the premises. Mental models theory suggests that the quality of reasoning improves for invalid unbelievable arguments due to a more elaborate search for alternative models. When an argument is valid, there are no inconsistent models. A more elaborate search will therefore produce no additional benefit.

Selective processing theory (SPT; Evans, 2007; Evans, Handley, & Harper, 2001; Klauer et al., 2000; Stupple, Ball, Evans, & Kamal-Smith, 2011) also describes the reasoning process as involving the construction of mental models but assumes that people normally attempt to construct only a single model of the premises (Evans, Handley, Harper, & Johnson-Laird, 1999), one that confirms a believable conclusion or disconfirms an unbelievable conclusion. Given that for valid syllogisms an inconsistent model of the premises can never be found, the acceptance rates of believable and unbelievable valid conclusions should be similar. With invalid conclusions, on the other hand, successful model creation leads to acceptance of believable conclusions but rejection of unbelievable conclusions. The possibility of finding a confirmatory model even when the conclusion is invalid can lead to a greater acceptance rate for believable conclusions.

SDT and the Response Bias Account

Support for reasoning-based accounts has centered on the interaction between validity and conclusion believability in endorsement rate analysis. Consistent with predictions, the difference in acceptance rates of valid and invalid arguments, hits (H) — false alarms (FA), is typically smaller for believable than for unbelievable conclusions. Dube et al. (2010) pointed out that SDT can account for this interaction purely in terms of response bias. In the standard SDT model, valid and invalid arguments are represented by normal distributions of evidence. A person chooses a response criterion, and H and FA rates are determined by the proportions of the distributions that exceed the criterion. When accuracy is constant (signifying no change in the quality of reasoning), moving the criterion will change H and FA rates. Plotting these changes produces a curvilinear receiver operating characteristic (ROC; see Figure 1) of the kind typically observed in empirical data (Dube, Rotello, & Heit, 2011). The nonlinear relationship between H and FA makes it possible for the H — FA difference to change in size despite no change in the accuracy of reasoning. Dube et al. noted that this can in theory account for the Validity × Believability interactions reported in the literature. ROC analysis of their own syllogism data showed that conclusion believability affected response bias but not accuracy.

The response bias account offers an appealingly simple alternative to extant theories of belief bias. In the present study, we applied SDT to syllogistic reasoning with two goals in mind. First, the fact that Dube et al.’s (2010) conclusions contradict so much earlier work makes it important to replicate and generalize their findings. Dube et al., like many previous studies, used a fixed set of problems for each condition. Such a design increases the danger that aspects of the materials other than believability will co-vary with argument validity. In attempting to replicate Dube et al.’s findings, we introduced a greater degree of randomization in our materials. In addition, we added two new tests of the response bias hypothesis. Experiment 1 used simple and complex syllogisms in order to examine the effect of logical complexity on belief bias. Experiment 2 examined the ability of cognitive capacity and time pressure to mediate the belief bias effect. The reasoning-based accounts described earlier predict that the influence of believability on reasoning should depend on all of these factors. The response bias account, on the other hand, predicts no effect of believability on the accuracy of syllogistic reasoning regardless of these manipulations.

The application of SDT to reasoning is of interest beyond Dube et al.’s (2010) specific hypothesis. Serious theoretical problems with the H — FA index (Dube et al., 2010; Klauer et al., 2000) make it difficult to interpret much of what has been published on belief bias. The second goal of our study was to extend the use of the SDT model in testing reasoning-based accounts. These accounts predict that believability can affect reasoning accuracy. However, the accuracy effect should be mediated by factors that influence the use of reasoning strategies such as logical complexity, cognitive capacity, and time pressure.
Experiment 1

Syllogism complexity was manipulated by presenting people with simple or complex syllogisms. The former are easy to solve because their proofs require fewer steps to determine the logical validity compared to the latter, which require a larger amount of steps and more difficult logical operations (e.g., reductio ad absurdum) in their proofs. In mental model terms, the simple syllogisms are easy to solve because they only allow for the construction of a single mental representation of the premises. Complex syllogisms, on the other hand, are difficult because up to three mental models are available. According to mental model based theories, in order to be certain that a multiple-model syllogism is valid, one needs to verify that the conclusion follows for all models of the premises. Previous research comparing simple and complex syllogisms has found that the logic by belief interaction is usually absent for the simple problems (Klauer et al., 2000; Newstead et al., 2002: although see also Gilinsky & Judd, 1994). The disappearance of the interaction has been put forward as evidence that belief does not have the same effect on reasoning in simple problems. However, in the SDT model the lack of an interaction is not incompatible with a pure response bias. According to the bias account, the content of the materials rather than their logical structure is the critical factor. It therefore predicts only response bias in both simple and complex syllogisms. The various reasoning-based accounts make predictions about the effect of belief in simple syllogisms that can also be tested by the SDT model. We now outline the predictions for the four dominant belief bias theories in terms of the SDT framework.

Selective scrutiny predicts more accurate reasoning for unbelievable compared to believable syllogisms because people tend to automatically accept believable conclusions but engage in deductive reasoning for unbelievable ones. This should be true for both simple and complex problems. Misinterpreted necessity predicts an accuracy effect for complex syllogisms because for conclusions that are indeterminately invalid, people adopt the heuristic of responding solely based on believability. This means that they will tend to be accurate when invalid conclusions are unbelievable but inaccurate when they are believable. This difference in accuracy is not predicted for simple syllogisms because they are never indeterminately invalid.

Mental models theory and selective processing theory both predict an accuracy effect for multiple-model (i.e., complex) syllogisms. According to MMT, unbelievable conclusions cue a motivated search for all possible models, making it likely that an invalid argument will be correctly rejected. Believable conclusions do not cue a search, so that believable invalid arguments are less likely to be rejected. According to SPT, unbelievable conclusions cue a falsifying strategy so that the single model constructed for an unbelievable invalid argument points to its correct rejection. On the other hand, believable conclusions cue a confirming strategy so that the single model constructed for a believable invalid argument points to its incorrect acceptance. The effect on accurate responding described by MMT and SPT does not apply to valid conclusions where there is only a single possible model (i.e., simple syllogisms). The net result is that with complex problems, both MMT and SPT predict an effect of believability on accuracy as a result of differences in the ability to reject invalid conclusions.

With simple problems, on the other hand, according to MMT people will always arrive at the same model regardless of differences in reasoning style or strategy. According to SPT, people will arrive at a confirmatory model for valid syllogisms with believable conclusions (leading to correct acceptance) but fail to arrive at a disconfirmatory model for valid syllogisms with unbelievable conclusions (also leading to correct acceptance). Thus, according to both theories, believability is predicted to have no effect on accuracy.

Method

Participants. Ninety-one undergraduate psychology students from Plymouth University participated for course credit.

Design. Logical validity (valid vs. invalid argument) and conclusion believability (believable vs. unbelievable) were manipulated within subjects, and argument complexity (simple vs. complex syllogisms) was manipulated between subjects.

Materials. For each participant, we created a unique list of 64 syllogisms (32 valid and 32 invalid) by randomly assigning 64 item contents to the available syllogistic structures. This ensured that item content had the potential to be different in every Validity × Believability cell for each participant, allowing us to control for content effects. Every list contained 32 valid syllogisms and 32 invalid syllogisms. The simple syllogisms were of the one-model type using all four figures and all quantifiers except the “Some . . . not” one. The complex syllogisms were of the multiple-model type and were taken from Dube et al. (2010). For a detailed explanation on the meaning of syllogistic figure, see Evans et al. (1999).

The conclusions were definitionally true or false by combining object categories with category members. In the simple argument condition, conclusions used the “No”- or “Some”-quantifiers. Believable “No”-conclusions consisted of a category with a member from a different category (e.g., no tools are trout). Unbelievable ones consisted of a category with one of its members (e.g., no tools are hammers). Believable “Some”-conclusions consisted of a category with one of its members (e.g., some tools are trout). Unbelievable ones consisted of a category with one of its non-members (e.g., some tools are hammers). Unbelievable conclusions featured the category followed by one of its members (e.g., some tools are hammers). Unbelievable conclusions featured a category member followed by its category (e.g., some hammers are not tools). We used nonsensical middle terms to control for premise believability. Each list was generated by combining the structures, item contents, and middle terms according to the constraints outlined above (see Table 1).

Procedure. Participants were tested on individual computers in small groups. The instructions were taken from Dube et al. (2010, Experiment 2) and were presented on the screen. Participants were instructed to assume the premises were true, to judge

---

1 Although reasoning-based theories focus on the effect of believability on invalid arguments, additional factors could increase the acceptance rate of valid arguments (e.g., Evans et al., 2001; Klauer et al., 2000; Newstead et al., 1992; Oakhill et al., 1989). One could imagine a simultaneous increase in H and FA rates that resembles a pure bias shift. Whether this is possible or likely depends on the nature of these additional factors. This does not alter the critical point that a change in accuracy due to believability serves to distinguish between the reasoning-based and pure bias accounts.
whether the conclusion necessarily followed, and to rate their confidence.

After completing four practice trials, the participants solved the remaining 64 syllogisms. Syllogisms were presented one at a time with response options (s = Valid, k = Invalid) shown at the bottom of the screen. After each validity judgment, participants indicated their confidence on a scale ranging from 1 (not confident) to 3 (very confident).

Results

Proportion measures (endorsement rates and $A_c$) were arcsine transformed to conform with the assumptions of analysis of variance (ANOVA). The simple and complex problem conditions were analyzed separately. This was justified by a preliminary analysis that revealed significant three-way interactions between validity, believability, and complexity in endorsement rates, $F(1, 89) = 9.83, p = .002, \eta_p^2 = .10$, and a significant Believability $\times$ Complexity interaction in accuracy, $F(1, 89) = 5.51, p = .021, \eta_p^2 = .058$.

Endorsement rates. To verify whether the participants showed the traditional belief bias effect, endorsement rates (see Table 2) were submitted to a 2 (logical status: valid vs. invalid) $\times$ 2 (conclusion believability: believable vs. unbelievable) repeated measures ANOVA.

Table 2

<table>
<thead>
<tr>
<th>Syllogism condition</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>Simple</td>
<td>.90 (.02)</td>
<td>.87 (.03)</td>
</tr>
<tr>
<td>Complex</td>
<td>.78 (.03)</td>
<td>.75 (.04)</td>
</tr>
</tbody>
</table>

Note. Standard errors of the per condition analyses are in parentheses.

Simple. Participants accepted more valid than invalid arguments, $F(1, 46) = 352.33, p < .001, \eta_p^2 = .89$. Participants also responded “valid” more to believable than to unbelievable conclusions, $F(1, 46) = 5.26, p = .026, \eta_p^2 = .10$. The two factors did not interact, $F < 1, p > .46$.

Complex. A main effect of logical status indicated that the participants on average made more “valid” responses for valid arguments compared to invalid arguments, $F(1, 43) = 74.14, p < .001, \eta_p^2 = .63$. A main effect of believability showed that people responded “valid” more to arguments with believable conclusions compared to arguments with unbelievable conclusions, $F(1, 43) = 5.45, p = .024, \eta_p^2 = .11$. Validity and believability interacted; the validity effect was larger for unbelievable than for believable arguments, $F(1, 43) = 10.22, p = .003, \eta_p^2 = .19$.

SDT analysis. ROCs (see Figure 2) were constructed from confidence ratings for each participant and conclusion type (believable or unbelievable). From these were derived measures (see Table 3) of accuracy, $A_c$ (an estimate of the area under the ROC), and bias, $c_a$ (more negative values indicate a more liberal response criterion).$^2$

Simple. An effect of believability on $c_a$ reflected more liberal responding to believable than to unbelievable conclusions, $F(1, 46) = 6.04, p = .018, \eta_p^2 = .12$. Believability did not affect accuracy, $F < 1, p > .73$.

Complex. An effect of believability on $c_a$ reflected more liberal responding to believable than to unbelievable conclusions, $F(1, 43) = 4.92, p = .032, \eta_p^2 = .10$. An effect of believability on accuracy indicated better reasoning for arguments with unbelievable conclusions than for arguments with believable conclusions, $F(1, 43) = 7.02, p = .011, \eta_p^2 = .14$.

$^2$ Rather than fit individual data, Dube et al. (2010) analyzed aggregate ROCs. Performing the same aggregate analysis on our own data did not change the overall conclusions reached (see the Appendix).
Discussion

Dube et al. (2010) suggested that the increased tendency to endorse believable conclusions is a bias that occurs at the response stage. If this is the case, then the believability of conclusions should have no effect on judgment accuracy irrespective of the complexity of the syllogisms. With simple syllogisms, the believability effect was consistent with a pure response bias. With complex syllogisms, however, unbelievable conclusions produced more accurate judgments. The latter finding is at odds with that of Dube et al., who found no effect of believability on accuracy with multiple-model syllogisms. Although differences in materials may be behind our divergent findings, individual differences suggest a more interesting possibility.

Several of the reasoning-based accounts (selective scrutiny, MMT, and SPT) suggest that people engage in more effortful or difficult reasoning when faced with unbelievable conclusions. If true, any factor that reduces the ability to engage in such reasoning should reduce the effect of believability on reasoning accuracy. Factors could include cognitive capacity, reasoning expertise, motivation, or situational constraints. For instance, previous belief bias research in individual differences has shown that cognitive capacity is negatively correlated with belief bias (Sá, West, & Stanovich, 1999). In Experiment 2, we considered two factors that have previously been associated with a disappearance of the Believability × Validity interaction: cognitive ability (Newstead, Handley, Harley, Wright, & Farrelly, 2004) and time constraints (Evans & Curtis-Holmes, 2005). General cognitive ability or general intelligence is correlated with working memory (Conway, Kane, & Engle, 2003), which is known to be associated with reasoning ability (Capon, Handley, & Dennis, 2003). In Experiment 2, participants were divided into high and low cognitive ability groups using a test of general intelligence. In addition, some participants were allowed to make judgments at their own pace, whereas others were given a 10-s response deadline. Reasoning-based accounts predict that both low cognitive ability and time pressure should reduce the effect of believability on accuracy. Such a demonstration would establish that differences in reasoning ability may lead some groups and not others to display accuracy effects.

Experiment 2

Method

Participants. Eighty-five undergraduate psychology students from Plymouth University participated for course credit.

Design. Logical validity (valid vs. invalid argument) and conclusion believability (believable vs. unbelievable) were manipulated within subjects. The time allowed to respond (10 s vs. no limit) and cognitive ability (high vs. low) were between-subjects factors.

Materials and measures. Problems were complex syllogisms generated in the same way as those presented in the complex condition of Experiment 1. Participants were given Part I of the AH4 Group Test of General Intelligence (Heim, 1970), which contains 65 verbal or numerical questions. Newstead et al. (2004) have shown that scores on Part I of the AH4 are related to logical performance on a variety of deductive reasoning tasks.

Procedure. The procedure was identical to that of Experiment 1 with the following exceptions. Participants first completed the AH4. Participants in the time limit condition were instructed to respond to the reasoning problems within 10 s while remaining as accurate as possible. A red bar at the top of the screen shortened as time ran out for each timed trial.

Table 3

Experiment 1: Signal Detection Theory Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Simple problems</th>
<th>Complex problems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>Accuracy ($A_1$)</td>
<td>.89 (.02)</td>
<td>.90 (.02)</td>
</tr>
<tr>
<td>Bias ($c_a$)</td>
<td>−0.01 (0.04)</td>
<td>0.18 (0.05)</td>
</tr>
</tbody>
</table>

Note. Lower $c_a$ values indicate a more liberal response criterion. Standard errors of the per condition analyses are in parentheses.
Results

The majority of responses (99%) were made within the time limit of 10 s. We divided the sample into high and low cognitive ability by performing a median split on Part I AH4 scores. In the unlimited time condition, 23 of 49 participants scored above the median of 38; the remaining participants scored at or below the median. In the limited time condition, 18 of 36 participants scored above the median. High and low cognitive ability groups were analyzed separately. This was justified by a preliminary analyses that revealed a marginally significant three-way interaction between believability, time limit, and cognitive ability for accuracy, $F(1, 81) = 3.22, p = .076, \eta^2_p = .09$. However, there was no significant Validity $\times$ Believability $\times$ Time Limit $\times$ Cognitive Ability interaction for endorsement rates, $F < 1, p = .41$. For this reason, the endorsement rate analysis is reported collapsed over cognitive ability.

Endorsement rates. Endorsement rates (see Table 4) were submitted to a 2 (logical status: valid vs. invalid) $\times$ 2 (time limit: 10 s vs. no limit) mixed ANOVA with logical status and believability as within-subjects factors and time limit as a between-subjects factor. Valid arguments were endorsed more than invalid arguments, $F(1, 83) = 64.1, p < .001, \eta^2_p = .44$. Participants responded “valid” more often in the untimed than in the timed condition, $F(1, 83) = 4.3, p = .042, \eta^2_p = .05$. Validity interacted with time limit condition such that the validity effect was larger in the untimed than in the timed condition, $F(1, 83) = 14.6, p < .001, \eta^2_p = .15$. Believable arguments were endorsed more than unbelievable arguments, $F(1, 83) = 15.9, p < .001, \eta^2_p = .16$. Validity, believability and time limit interacted such that the interaction index was positive in the untimed (index = 0.15) condition and negative in the timed condition (index = −0.26), $F(1, 83) = 15.1, p < .001, \eta^2_p = .15$.

SDT analysis. ROCs (see Figure 3) produced measures of accuracy ($A$) and bias ($c_a$) (see Table 5), which were each submitted to a 2 (believable vs. unbelievable) $\times$ 2 (10-s limit vs. no limit) mixed ANOVA.

High cognitive ability. There was a marginal effect of time limit, with higher performance in the unlimited than the limited time condition, $F(1, 39) = 3.47, p = .07, \eta^2_p = .08$. There was no main effect of believability ($p > .44$), but time limit and believability interacted, $F(1, 39) = 12.36, p = .001, \eta^2_p = .24$. Post hoc tests revealed that in the unlimited time condition, participants were more accurate for unbelievable than for believable conclusions, $t(22) = 2.46, p = .022$. In the time limit condition, performance was worse for unbelievable conclusions than for believable conclusions, $t(17) = 2.44, p = .026$. Analysis of $c_a$ revealed that participants were more liberal in the unlimited time compared to the limited time condition, $F(1, 39) = 5.65, p = .022, \eta^2_p = .13$. Participants responded more liberally to believable than to unbelievable conclusions, $F(1, 39) = 12.36, p = .001, \eta^2_p = .24$. Time limit and believability did not interact ($p > .79$).

Low cognitive ability. Participants with unlimited time reasoned more accurately than participants with limited time, $F(1, 42) = 5.22, p = .027, \eta^2_p = .11$. No other effects were significant ($ps > .46$). Analysis of $c_a$ revealed that participants responded more liberally to believable than to unbelievable conclusions, $F(1, 42) = 5.71, p = .021, \eta^2_p = .12$. No other effects were significant ($ps > .42$).

Discussion

According to reasoning-based accounts, the effect of believability on reasoning accuracy should be most pronounced when conditions are conducive to formal reasoning. Our findings were consistent with this prediction. In the unlimited time condition, only the high cognitive ability group was more accurate when judging unbelievable compared to believable conclusions. The low cognitive ability group showed only a liberal response bias to believable conclusions, a pattern like that reported by Dube et al. (2010). The latter was not due to floor performance as accuracy of the low ability group remained well above chance. Although Experiment 2 was designed to examine the relationship between cognitive capacity and reasoning strategies, we conducted a final analysis to allow a direct comparison with the complex condition of Experiment 1. Collapsing over the ability groups in the unlimited time condition revealed a marginally significant effect of believability on accuracy, $F(1, 48) = 3.46, p = .069, \eta^2_p = .07$. The contrast between this weak effect in Experiment 2 and the robust one observed in Experiment 1 would be puzzling had we not considered the possibility that different subgroups within a sample might respond differently to believability. The finding of a clear accuracy effect only within the subgroup predicted by reasoning-based accounts underscores the usefulness of the individual differences approach.

Also consistent with the prediction of reasoning-based accounts, improved accuracy for unbelievable conclusions was not observed when a response deadline limited the ability to engage in formal reasoning. Surprisingly, judgments were more accurate for believable conclusions (although this was significant only for the high ability group). The finding is not predicted by reasoning-based accounts and requires additional assumptions. Participants were apparently unable to judge unbelievable conclusions with better than chance accuracy given the short time available. Evans and Curtis-Holmes (2005) investigated the effect of time pressure on belief bias in syllogisms and found no difference in performance under time pressure between believable and unbelievable syllogisms. However, they did observe a larger drop in accuracy for unbelievable...

Table 4

<table>
<thead>
<tr>
<th>Group</th>
<th>Valid</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
<td>Believable</td>
<td>Unbelievable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High cognitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>capacity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time limit</td>
<td>.73 (.05)</td>
<td>.46 (.07)</td>
<td>.48 (.05)</td>
<td>.42 (.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No limit</td>
<td>.83 (.03)</td>
<td>.73 (.06)</td>
<td>.57 (.06)</td>
<td>.32 (.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low cognitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>capacity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time limit</td>
<td>.70 (.04)</td>
<td>.49 (.06)</td>
<td>.57 (.05)</td>
<td>.51 (.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No limit</td>
<td>.70 (.04)</td>
<td>.67 (.06)</td>
<td>.49 (.05)</td>
<td>.39 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. High cognitive capacity participants scored above the median Part I AH4 Group Test of General Intelligence score. Low cognitive capacity participants scored at or below the median. Standard errors for the per condition analyses are in parentheses.
able compared to believable conclusions, suggesting that time pressure interferes more with the former. This hints that with an increase in time pressure, they might have uncovered the sort of reversal found here.

A possible explanation for the reversal is that unbelievable conclusions are more difficult to initially process. Klauer et al. (2000) suggest, for example, that it may be particularly difficult to construct a coherent semantic model when the elements of a problem conflict with prior beliefs. This might appear at odds with findings that believable invalid conclusions produce the slowest judgments (Ball, Phillips, Wade, & Quayle, 2006; Stupple et al., 2011; Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003). However, overall response time includes not only the construction of a problem representation but also the decision process that follows. Believable invalid problems may be quickly processed initially but the conflict between belief and reasoning leads to a more drawn out decision (Thompson et al., 2003).

**General Discussion**

Reasoning-based accounts of belief bias have traditionally focused on the H-FA index. This index, aside from other problems (Dube et al., 2010; Klauer et al., 2000), is inadequate for making the theoretically critical distinction between changes in response bias and accuracy. SDT offers a formal way to model this distinction. Dube et al. (2010) found that ROC analysis of their own data showed conclusion believability to affect only response bias and not accuracy. They also pointed out that, in theory, the majority of previous findings could also be explained solely in terms of response bias, obviating the need to postulate changes in the quality of reasoning. We found that participants were more accurate when judging unbelievable compared to believable conclusions. However, the effect on accuracy was absent under some conditions. With simple syllogisms, with participants of lower cognitive ability, and under time pressure, there was no accuracy advantage for unbelievable conclusions. The liberal response bias...
for believable conclusions did persist under all of these conditions. It appears that response bias is a component of belief bias, but it is only one component (see Klauer & Kellen, 2011, for a similar conclusion).

The SDT analysis also allowed us to test a number of predictions made by the reasoning-based accounts. Selective scrutiny predicts an effect on accuracy regardless of logical complexity. There was no such effect for simple problems, only for complex problems. Misinterpreted necessity predicts this finding but does not explain the disappearance of the accuracy effect under time pressure or for participants with lower cognitive ability. Misinterpreted necessity attributes the accuracy effect to a failure in formal reasoning. The accuracy effect should therefore be exacerbated, not diminished, under conditions that make reasoning more difficult. In sum, our findings were inconsistent with both selective scrutiny and misinterpreted necessity.

MMT claims that the accuracy effect stems from a motivated search for alternative models for unbelievable conclusions. SPT predicts the accuracy effect on the basis of a falsifying model construction for unbelievable conclusions. These are both effortful operations that require sufficient cognitive effort and adequate time; as such our findings are compatible with both accounts, although the current experiments do not allow us to distinguish between them. These predictions of the reasoning-based theories differ in the strategies that they bring to a reasoning task. Previous research has identified a number of factors that mediate individual difference in the strategies that reasoners use (Klauser & Kellen, 2011). There is a growing awareness in the literature that individuals differ in the strategies that they bring to a reasoning task. Previous research has identified a number of factors that mediate individual difference in the strategies that reasoners use (Klauser & Kellen, 2011).

The results of Experiment 2 suggest that inter-sample differences could explain the discrepancy between our findings and those of Dube et al. (2010). It might seem puzzling that the overall performance of Dube et al.’s participants was similar to that of our participants in the conditions where an accuracy effect was observed. There are two things to consider. Cognitive ability may be one factor mediating the use of reasoning strategies, but there are likely other factors such as reasoning expertise, task familiarity, and motivation. It is conceivable that two samples differing in multiple factors could show comparable levels of overall reasoning but different tendencies in their use of strategies. In addition, although higher cognitive capacity seems to be associated with the use of reasoning strategies, it is likely true that the best reasoners ignore these strategies and apply optimal reasoning to all problems regardless of believability. If a sample is a mixture of poor, best, and “merely good” reasoners, with only the last group using strategies in which believability dictates the quality of reasoning, the relationship between overall accuracy and the emergence of an accuracy effect in the aggregate data is not straightforward and would depend on the relative proportions of these subgroups.

There is converging evidence for the existence of qualitatively different subgroups of reasoners from neuroimaging (Reverberi et al., 2012) and the examination of response times (Stipple et al., 2011). Other individual differences that have been linked to subgroup membership include working memory ability (Capon et al., 2003) and the ability to suppress intuitive responses (Frederick, 2005). Processing style factors such as the tendency to generate counterexamples (Newstead et al., 2002) and need for cognition (Stanovich & West, 1999) might predict subgroup membership in addition to the other factors discussed so far.

There is a growing awareness in the literature that individuals differ in the strategies that they bring to a reasoning task. Previous research has identified a number of factors that mediate individual difference in strategy. Experiment 2 examined one of these, cognitive capacity, and showed that people who scored higher or lower on a test of general intelligence seemed to respond in qualitatively different ways to conclusion believability. The implication is that if the proportion of different subgroups of reasoners varies between samples, performance may seem puzzlingly inconsistent. A clear story may only emerge by looking beyond aggregate data to examine performance at the level of subgroups. Importantly, our findings jeopardize the conclusions drawn from many previous belief bias experiments that considered only aggregate performance when examining the effect of various experimental manipulations such as complexity (e.g., Newstead et al., 1992), response time taken (e.g., Evans & Curtis-Holmes, 2005), and instructions (Evans, Newstead, Allen, & Pollard, 1994).

Table 5

<table>
<thead>
<tr>
<th>Group</th>
<th>Time limit</th>
<th>No time limit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>High cognitive capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (A)</td>
<td>0.68 (.03)</td>
<td>0.54 (.06)</td>
</tr>
<tr>
<td>Bias (c)</td>
<td>-0.28 (.12)</td>
<td>0.17 (.09)</td>
</tr>
<tr>
<td>Low cognitive capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (A)</td>
<td>0.56 (.03)</td>
<td>0.51 (.05)</td>
</tr>
<tr>
<td>Bias (c)</td>
<td>-0.31 (.08)</td>
<td>-0.01 (.12)</td>
</tr>
</tbody>
</table>

Note. High cognitive capacity participants scored above the median Part I AH4 Group Test of General Intelligence score. Low cognitive capacity participants scored at or below the median. Standard errors for the per condition analyses are in parentheses.
References


(Appendix follows)
Appendix

Aggregate Signal Detection Theory (SDT) Analyses for Experiments 1–2

To rule out the possibility that our findings differed from those of Dube, Rotello, and Heit (2010) because of our use of individual-level—rather than aggregate-level—parameter tests, we fit the SDT model described by Dube et al. to aggregate receiver operating characteristic data in each condition (see Table A1). The conclusions obtained from the aggregate analysis differed from the individual analysis in only one case: The pure-bias model was rejected in the Experiment 1 simple condition. However, it seemed that the extremely poor fit of both the constrained and the unconstrained models in this condition (both $G^2 > 150$) might be due to the number of reasoners performing at ceiling. We therefore performed a median split according to overall accuracy and examined the below-median group, who would not be at ceiling. The fit of the models was improved, and the pure-bias model was not rejected, $G_{\text{unconstrained}}^2 = 83.83$ – $G_{\text{constrained}}^2 = 79.93 = 3.9$, $df = 2$, $p = .14$.

Table A1

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>$G^2$ difference</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1: Complex</td>
<td>22.24 – 14.47 = 7.77</td>
<td>2</td>
<td>.021</td>
</tr>
<tr>
<td>Experiment 1: Simple</td>
<td>164.25 – 153.5 = 10.75</td>
<td>2</td>
<td>.005</td>
</tr>
<tr>
<td>Experiment 2: Untimed</td>
<td>34.20 – 30.12 = 4.08</td>
<td>2</td>
<td>.13</td>
</tr>
<tr>
<td>Experiment 2: Untimed + High Ability</td>
<td>19.52 – 13.23 = 6.29</td>
<td>2</td>
<td>.043</td>
</tr>
<tr>
<td>Experiment 2: Untimed + Low Ability</td>
<td>21.58 – 21.19 = 0.39</td>
<td>2</td>
<td>.82</td>
</tr>
<tr>
<td>Experiment 2: Timed</td>
<td>26.30 – 3.64 = 22.66</td>
<td>2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experiment 2: Timed + High Ability</td>
<td>24.95 – 4.03 = 20.94</td>
<td>2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experiment 2: Timed + Low Ability</td>
<td>11.22 – 6.64 = 4.58</td>
<td>2</td>
<td>.10</td>
</tr>
</tbody>
</table>

Note. $p < .05$ indicates a rejection of the null hypothesis that belief affects only response bias. The first $G^2$ value represents a constrained model in which $\mu_{\text{believable}} = \mu_{\text{unbelievable}}$ and $\sigma_{\text{believable}} = \sigma_{\text{unbelievable}}$. The second $G^2$ value represents a model in which the four parameters are unconstrained. The difference is compared to a $\chi^2$ distribution with 2 degrees of freedom.

Received May 28, 2012
Revision received February 5, 2013
Accepted February 12, 2013