Logic, Beliefs, and Instruction:
A Test of the Default Interventionist Account of Belief Bias

Simon J. Handley, Stephen E. Newstead, and Dries Trippas
University of Plymouth

According to dual-process accounts of thinking, belief-based responses on reasoning tasks are generated as default but can be intervened upon in favor of logical responding, given sufficient time, effort, or cognitive resource. In this article, we present the results of 5 experiments in which participants were instructed to evaluate the conclusions of logical arguments on the basis of either their logical validity or their believability. Contrary to the predictions arising from these accounts, the logical status of the presented conclusion had a greater impact on judgments concerning its believability than did the believability of the conclusion on judgments about whether it followed logically. This finding was observed when instructional set was presented as a between-participants factor (Experiment 1), when instruction was indicated prior to problem presentation by a cue (Experiment 2), and when the cue appeared simultaneously with conclusion presentation (Experiments 3 and 4). The finding also extended to a range of simple and more complex argument forms (Experiment 5). In these latter experiments, belief-based judgments took significantly longer than those made under logical instructions. We discuss the implications of these findings for default interventionist accounts of belief bias.

Keywords: reasoning, dual processes, belief bias, instruction, conditionals

The idea that human thinking is influenced by the operation of two systems of thought is now a widely accepted position across many subdisciplines of psychology. According to this view, judgments are often guided by rapid, unconscious associative processes, an idea that has been applied in a range of research areas, including learning, memory, reasoning, and social judgment (Evans, 2008). However, it is also recognized that some processes are slow, conscious, and deliberative, drawing on the resources of working memory and executive functions (Evans, 2003). This has led many theorists to argue that responses to reasoning and decision-making tasks arise from the operation of two distinct cognitive systems associated with qualitatively different processing characteristics. Type 1 processes (using Evans’s 2009 terminology) are rapid, automatic, preconscious, and relatively undemanding of computational capacity. In contrast, Type 2 processes are controlled, conscious, analytic, and related to individual differences in working memory capacity and general intelligence.

In the reasoning literature, researchers often categorize the responses that their participants generate into these two kinds, variously termed analytic or heuristic, deliberative or automatic, rational or intuitive, or rule-based or associative. In such categorizations, a fairly direct correspondence is assumed between a response and its underlying process, such that intuitive responses based upon an apparently superficial problem characteristic are assumed to reflect Type 1 processing, whereas analytic responses based upon the logical structure of an argument are assumed to reflect Type 2 processing. Consider, for example, an argument of the following kind in which there is a conflict between the believability of the conclusion and its underlying logic status (Sá, West, & Stanovich, 1999):

- All plants need water.
- Roses need water.
- Therefore, roses are plants.

According to dual-process accounts, the intuitive or Type 1 response is to endorse the conclusion because it is consistent with underlying beliefs, whereas the appropriate rejection of the conclusion is assumed to reflect more deliberative, analytic, or Type 2 processes (Evans, 2009). A key question concerns the way in which these processes interact with one another. A number of proposals in the literature can be broadly categorized as parallel processing or default interventionist accounts (see Evans, 2007b; 2008). Parallel processing models are common in social psychology (see Chaiken & Trope, 1999) but have also been applied in judgment and decision making through a distinction made by Sloman (1996) between rule-based and associative processes. These models assume that Type 1 and Type 2 processes are initiated in parallel, with Type 2 responses dominating if sufficient resources are available. An alternative and widely held view in the psychology of reasoning proposes that the many demonstrations of biases extant in the literature can be attributed to the dominance of Type 1 processes, which suggest a compelling, if erroneous, response to a problem before Type 2 processes are fully engaged (Evans, 2007a). In order to generate an alternative response, Type 2 processes must undertake the working memory–demanding task of inhibiting the Type 1 output (see, for example, Handley, Capon,
Beveridge, Dennis, & Evans, 2004) and produce an alternative representation of the premises (Evans, 2007a). These accounts have been labeled default interventionist (DI) because they assume that Type 1 processes cue a default response, which must then be resisted in favor of more deliberative processing. However, given that the initial answer is often compelling (Thompson, 2009) and generating the alternative requires cognitive effort (Stanovich, 2009), the initial answer is often accepted.

One common reasoning phenomenon that has been explained in such a manner is belief bias of the kind previously illustrated. Belief bias is ubiquitous and refers to the tendency of individuals to endorse conclusions on the basis of their believability rather than their validity, even when they have been instructed to set aside beliefs and draw only logically necessary inferences (Evans, Barston, & Pollard, 1983). Stanovich (1999), for example, viewed belief bias as an instance of the fundamental computational bias, that is, the tendency to reason on the basis of highly personal, contextualized representations as opposed to decontextualized ones. Researchers have attempted (with modest success) to ameliorate belief bias by use of strong instructions to reason logically under the assumption that this will cue the reasoner to engage in Type 2 thinking (Evans, Newstead, Allen, & Pollard, 1994; Evans, Handley, Neiins, & Over, 2009). There is some evidence that belief bias is more common amongst reasoners of lower cognitive capacity (see, for example, Stanovich, 1999), suggesting that those who lack the resources to reformulate and maintain an alternative representation in working memory are more likely to rely on the output of Type 1 processes. There is also some evidence that belief bias on certain tasks (syllogistic reasoning) is more likely when individuals are reasoning under time pressure (see, for example, Evans & Curtis-Holmes, 2005) or under a secondary working memory load (De Neys, 2006), which has been interpreted as supporting the primacy of belief-based processes.

Despite a large body of evidence suggesting that belief bias is a typical manifestation of a fast, automatic Type 1 process, several pieces of recent evidence suggest a more complex picture. Specifically, it appears that belief bias in conditional reasoning, unlike syllogistic reasoning, is unaffected by speeded tasks (Evans, Handley, & Bacon, 2009); measures of belief bias in syllogistic reasoning do not consistently correlate with measures of cognitive capacity (Newstead, Handley, Harley, Wright, & Farrelly, 2004); and belief bias among primary-school-age children is positively rather than negatively correlated with working memory capacity (Morsanyi & Handley, 2008). Recent work on everyday conditional reasoning has shown that participants with greater cognitive capacity are more able to retrieve and selectively use counterexamples in rejecting invalid conditional inferences (Verschuereen, Schaeeken, & d’Ydewalle, 2005). The capacity to selectively retrieve relevant counterexamples to invalid inferences while at the same time inhibiting relations that could undermine valid inferences is interfered with by secondary tasks (De Neys, Schaeeken, & d’Ydewalle, 2005). The capacity to use knowledge in this way has also been shown to increase with development (Janveau-Brennan & Markovits, 1999). These findings suggest that, at the very least, belief effects in reasoning are not wholly dependent upon Type 1 processes.

There are, moreover, at least two pieces of evidence indicating that outcomes normally attributed to Type 2 processes, namely drawing inferences on the basis of a logical form of an argument, may be automatic and fast rather than deliberate and slow. Rader and Sloutsky (2002) presented participants with a conditional statement and the minor premise to either the modus ponens (MP) or affirmation of the consequent (AC) inference (e.g., if the weather is nice, Ed takes a walk; the weather is nice [MP] or Ed takes a walk [AC]) embedded in a story context. Relative to a control condition (e.g., Ed wondered whether the weather was nice), participants were faster to recognize the conclusions to the inferences, even though it was not necessary to draw the inference to understand the story. The authors concluded that these inferences are drawn automatically in the process of understanding the conditional statement. In a similar vein, Lea and her colleagues (Lea, 1995; Lea, O’Brien, Fisch, Noveck, & Braine, 1990) have shown, using a semantic priming technique, that simple “or-elimination” inferences (p or q, not p, therefore q) are drawn routinely in a text-processing task when the premises are available, even when the inference is not required to maintain coherence. Similar findings have been reported on transitive inference problems, where awareness does not appear to be a necessary condition for logical inferences to be drawn (Leo & Greene, 2008). The evidence that people draw simple inferences automatically when premise information is readily available suggests that, at least in some cases, logical reasoning may be accomplished by Type 1, rather than by Type 2, processes.

Together these findings suggest that reasoning based upon beliefs can often draw upon effortless processing, whereas reasoning based upon the underlying logical structure of an argument can often be accomplished fairly automatically in the process of comprehension. Given this analysis, why is it so commonly assumed that belief-based reasoning is automatic and logical reasoning effortful? In order to address this question, one must consider the way in which reasoning tasks are ordinarily administered. Consider, for example, the following example of a problem in which logic and belief are in conflict:

All mammals have four legs.
Whales are mammals.

Therefore, all whales have four legs.

Typically participants are given logical instructions that specify that they should (a) assume that the premises are true and (b) endorse only conclusions that necessarily follow. Participants who are more able to resist the influence of beliefs under these conditions tend to have greater cognitive capacity (Sá et al., 1999), and belief-based responses increase under conditions of a secondary task (De Neys, 2006) or speeded task (Evans & Curtis-Holmes, 2005). As we have seen, many authors have argued that these findings support DI accounts of the task where the inhibition of a belief-based response in favor of a logical one requires sufficient processing resources. However, the findings may have nothing to do with the status of different responses in terms of the processing systems underlying them; they may instead reflect the demands of applying a specific instructional set. If this conjecture is right, then belief-based responding does not have to depend upon Type 1 processing but may simply arise because of a failure to apply the task instructions appropriately.

Recently, a number of researchers have begun to examine performance on reasoning tasks under different instructional sets. For example, Heit and Rotello (2010; see also, Rotello & Heit, 2009)
have shown that when participants are instructed to make judgments of argument strength, they are more affected by factors such as similarity and argument length, whereas when participants are instructed to make deductive judgments, they are more affected by argument validity. In a similar way, Rips (2001) manipulated the causal consistency and logical necessity of conclusions under instructions to judge causal strength or argument validity. Rips reported a larger effect of argument validity than causal consistency on deduction judgments and a reversed pattern for judgments of causal strength. These findings suggest that inductive and deductive judgments are accomplished through distinct sets of processes, and the findings are inconsistent with the view that both deductive and inductive judgments are made with reference to a single probabilistic dimension (see, for example, Oaksford & Chater, 2001). This work is important in illustrating the value of instructional manipulations for discriminating between competing accounts of human reasoning, a strategy that we have adopted here in order to test the specific predictions of DI accounts of belief bias.

We will present a series of studies that were designed to examine the relative difficulty of generating a logical or a belief-based response to belief–logic conflict problems. A key characteristic of these problems is that there is a conflict between the response cued by beliefs and the response based on the underlying logical structure. In the experiments that follow, participants were presented with a series of problems of this kind and were required to either (a) give a logical response or (b) give a response based upon beliefs. Our principal aim was to compare accuracy of response and the time course of processing under each instructional condition. If the DI account is the right one, then we would expect belief-based judgments, which are assumed to be based upon default responses, to be both more accurate and take less time than judgments concerning the logical status of the conclusion, predictions that we flesh out more fully in the next section. Alternatively, if we are right in our conjecture that logical reasoning can be accomplished relatively automatically and belief-based responding is often effortful; the effect of instructions would be quite different. If this is the case, contrary to these accounts, judgments concerning the believability of conclusions should be less accurate and more prone to interference of logical validity and should take more time than judgments of logical validity.

**Experiment 1**

In Experiment 1, participants were presented with a series of MP conditional syllogisms based upon premises that were either factual or contrary to fact (see Table 1). An advantage of using such simple arguments is that it is easy to generate items in which the logical and belief-based responses conflict with one another (see Argument A in Table 1) or coincide (see Argument B), which allows an evaluation of the extent to which belief–logic conflict impacts upon logical or belief-based judgments. Of course, contrary-to-fact arguments, such as Argument A, require a “yes” response under logical instructions but a “no” response under belief-based instructions. However, it is relatively easy to control for conclusion polarity under different instructional sets by presenting the opposite to the logical conclusion for half of the problem sets (see Argument C in Table 1). The same manipulation can be introduced for factual problems (see Argument D in Table 1). This ensures that an equal number of belief-based and logical responses require yes and no conclusions across the problem set as a whole. In addition, using antonym pairs in the consequent clause for each problem (e.g., happy/sad) allows any possible confounding effect of explicit negation on response rates for different argument types to be controlled.

In Experiment 1, we sought to examine the impact of belief–logic conflict on accuracy and latency of responding for participants who were instructed to respond either logically or on the basis of their beliefs. The major question of interest was whether belief–logic conflict would have its greatest impact on judgments of conclusion validity or on judgments of conclusion believability. As we have shown, DI accounts emphasize the primacy of beliefs, which, when in conflict with logic, must be inhibited in favor of more effortful processing in order for a logical response to be generated. It is important to note that the beliefs that we are referring to here are of the kind commonly used in studies of belief bias in human reasoning and are of the kind presented earlier. These sorts of beliefs are deterministic, given that they are based upon empirical or definitional truths. Under DI accounts, reasoning on the basis of these sorts of beliefs depends upon default processing and should be rapid and relatively unaffected by the logical structure of the problem. On the basis of this analysis, the following dual-process predictions were derived:

**DP1:** Judgments concerning conclusion believability will be quicker than judgments of conclusion validity. This prediction follows readily from DI accounts of belief bias, where the primacy of belief-based processing (see Stanovich, Toplak, & West, 2008) underlies predictions concerning the impact of speeded task or secondary load on patterns of performance (see De Neys, 2006; Evans and Curtis-Holmes, 2005).

**DP2:** Belief judgments will result in fewer errors than logic judgments. This prediction follows from one of the principal assumptions of DI accounts: that Type I associative processes are rapid, automatic, and relatively effort free (Stanovich, 1999).

---

### Table 1

<table>
<thead>
<tr>
<th>Conflict problem</th>
<th>Nonconflict problem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Argument A</strong></td>
<td><strong>Argument B</strong></td>
</tr>
<tr>
<td>If a child is crying, then he is happy.</td>
<td>If a child is crying, then he is sad.</td>
</tr>
<tr>
<td>Suppose a child is crying.</td>
<td>Suppose a child is crying.</td>
</tr>
<tr>
<td>Is the child happy?</td>
<td>Is the child sad?</td>
</tr>
<tr>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
</tr>
<tr>
<td>Logic: YES</td>
<td>Logic: YES</td>
</tr>
<tr>
<td>Belief: NO</td>
<td>Belief: YES</td>
</tr>
<tr>
<td><strong>Argument C</strong></td>
<td><strong>Argument D</strong></td>
</tr>
<tr>
<td>If a child is crying, then he is happy.</td>
<td>If a child is crying, then he is sad.</td>
</tr>
<tr>
<td>Suppose a child is crying.</td>
<td>Suppose a child is crying.</td>
</tr>
<tr>
<td>Is the child sad?</td>
<td>Is the child happy?</td>
</tr>
<tr>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
</tr>
<tr>
<td>Logic: NO</td>
<td>Logic: NO</td>
</tr>
<tr>
<td>Belief: YES</td>
<td>Belief: NO</td>
</tr>
</tbody>
</table>
**DP3:** Belief–logic conflict will have a greater impact on logic judgments than belief judgments. This prediction follows from the DI assumption that belief-based responses are generated by default and consequently available prior to the onset of logical processing (see, for example, Evans, 2007b).

**Method**

**Design.** Participants were presented twice with a set of 40 conditional reasoning problems, all of which took the MP logical form. Half of the problems were presented accompanied by logical instructions, and the other half were presented with belief-based instructions. Half of the participants received the 40 arguments with logical instructions first, followed by the same 40 arguments under belief instructions. The remaining participants received the problems in the reverse order. Within each problem block, there were 20 conflict problems and 20 nonconflict problems, half of which required a yes response and half of which required a no response, resulting in 10 problems in each cell of the 2 (problem type) × 2 (response type) × 2 (instructional set) repeated-measures design. We analyzed performance on the first block of problems separately to evaluate the effect of instructional set as a between-participants variable.

**Participants.** The participants were second-year undergraduate students at the University of Plymouth who participated in the study as part of a course requirement. The 23 women and 7 men were randomly allocated to one of the two instructional conditions for the first block of 40 problems. Fifteen participants completed the first block of 40 problems under belief instructions and then the second block under logic instructions. The remaining 15 participants received the problems in the reverse order.

**Materials.** Table 1 shows an example of one of the problem contents used in Experiment 1. Each type of problem content was used to generate one of the four problem types shown in Table 1. The consequent clause in each factual conditional contained an adjective for which there was an available antonym that could be used to generate a contrary-to-fact version. The conclusion to each argument was either consistent with the logical conclusion to the MP argument or the opposite of the logical conclusion, ensuring that there was an equal number of logically valid and logically invalid conclusions and an equal number of believable and unbelievable conclusions. Half of the problems, the conflict problems, were based upon contrary-to-fact premises, in which there was a conflict between logic and belief, and half of the problems, the nonconflict problems, were based upon factual premises, in which there was no conflict between logic and belief. A total of 40 problems were generated, 20 conflict and 20 nonconflict problems, half of which had logically valid conclusions and half of which had logically invalid conclusions. Half of the conclusions were believable, and half were unbelievable. The 40 problems were presented under both logical and belief based instructions.

**Procedure.** Participants were told that they would be presented with two sets of 40 reasoning problems and that they would be required to respond to the first set of problems either on the basis of their beliefs or according to what logically followed. They were then presented with instructions on a computer screen for each of the instructional conditions as follows:

When instructed to answer according to your beliefs, this means you must answer according to your knowledge of what you know to be true in the world.

For example: If you finish a drink, then the glass will be full.

Suppose you finish your drink.

Will your drink be full?

The belief-based answer is NO, because based upon our knowledge of the world, we know that if a drink is finished, then the glass will be empty.

Participants were then presented with a second screen that displayed the logical instructions:

In contrast, when instructed to answer according to logic, this means you must assume each premise is true (even if in reality it is not true) and respond with the answer which logically follows from the statements presented.

For example: If you finish a drink, then the drink will be full.

Suppose that you finish your drink.

Will your drink be empty?

The logic-based answer to this problem is NO, because you must assume that if a drink is finished, then it will be full; therefore, the glass will not be empty.

Participants were then presented with a third instructional screen, indicating whether they should respond on the basis of their beliefs or on the basis of logic to the first block of 40 problems. Each problem was presented in full with the premises and the conclusion clearly separated. The words YES and NO appeared below the conclusion of each problem, and participants were instructed to press the YES button on the keyboard if the conclusion was believable (or followed logically) and NO if the conclusion was unbelievable (did not follow logically). Latency of response was recorded from the presentation of the full problem on the screen until a response was provided. The first and second blocks of problems were separated by another instructional screen indicating that participants should respond to the second set of problems on the basis of beliefs (if the first block was logic) or logic (if the first block was beliefs).

**Results and Discussion**

**Data treatment.** The analysis was completed in two stages. We began by examining performance on the first block of problems, where half of the participants received logical instructions and half of the participants received belief instructions. This allowed us to evaluate the effect of instructional set as a between-participants variable. We also analyzed overall performance across both problem blocks where instructional set was a within-subjects variable.

Prior to analyzing the data, we examined individual performance under each instructional condition in order to eliminate participants who showed evidence of failing to engage with the task instructions. Under both instructional sets, the nonconflict items required the same response. In contrast, the conflict items required a different response under belief and logic instructions. Consequently, performance on these problems was used as an indicator of task engagement. We chose to eliminate participants...
who scored below chance on these problems, as this provided some indication that an inappropriate instructional set was applied. However, it could be argued this is quite an arbitrary judgment, because poor performance can just as easily arise because conflict problems are difficult, given that there is more than one basis on which a response can be generated. If it is the case that errors are more common under one instructional set than another, one can reduce differences between these conditions by eliminating low-scoring participants. Nevertheless, we judged that this was a reasonably conservative criterion that, if anything, would reduce any observed effects rather than magnify them. In addition, for all the significant findings, we report the statistics for both the reduced and the full samples, which in the majority of cases showed equivalent effects. Given the possibility of ceiling effects on these simple problems, particularly on nonconflict items, we transformed all accuracy data prior to analysis using an arcsine transformation. This transformation improves homogeneity of variance across conditions, mitigating the influence of ceiling effects (Millsigan, 1987). Percentage accuracy rates prior to transformation are observed effects rather than magnify them. In addition, for all the reasonable criterion that, if anything, would reduce any scoring participants. Nevertheless, we judged that this was a reasonable conservative criterion that, if anything, would reduce any observed effects rather than magnify them.

Analysis of response latencies was carried out only on latencies for trials in which a correct response was generated. Raw latencies were used throughout, and cells with missing data were replaced with the sample mean for the relevant condition. In all of the experiments reported in this article, missing data accounted for fewer than 4% of the data points in any reported analysis.

In Experiment 1, three participants were eliminated from the analysis of the first block of data. All three participants showed below-chance accuracy on conflict problems under belief-based instructions. One additional participant was eliminated from the overall analysis, because he or she scored below chance on the conflict problems presented in the second block.

**Analysis of problem Block 1.** Table 2 shows the percentage accuracy and latency for correct responses under each instructional set in the first block of problems. A 2 (instruction) × 2 (problem type) × 2 (yes/no response) analysis of variance (ANOVA) on arcsine-transformed accuracy rates with repeated measures on the second two factors showed a main effect of instructions, $F(1, 25) = 4.41, p = .046, \eta^2_p = .15$; full sample $F(1, 28) = 7.85, p = .009, \eta^2_p = .22$, with greater accuracy under logical (99%) than under belief instructions (95%). There was also a main effect of problem type, $F(1, 25) = 23.67, p < .001, \eta^2_p = .49$; full sample $F(1, 28) = 18.46, p < .001, \eta^2_p = .40$, which showed, as expected, poorer performance on conflict than on nonconflict problems (94% vs. 99.8%, respectively). There was a significant interaction between these two factors, $F(1, 25) = 6.50, p = .017, \eta^2_p = .21$; full sample $F(1, 28) = 9.17, p = .005, \eta^2_p = .25$. Planned comparisons on this interaction showed a significant effect of problem type under belief instructions, $F(1, 25) = 24.71, p < .01, \eta^2_p = .50$, but no difference in performance between conflict and nonconflict problems under logic instructions, $F(1, 25) = 3.01, p = .09, \eta^2_p = .11$. This indicates that belief–logic conflict has more of an influence on judgments concerning the believability of a conclusion than judgments concerning its validity. There was no effect of response type ($F < 1$), and none of the remaining interactions were significant (all $F_s < 1$).

The equivalent ANOVA analysis on the latencies for correct responses showed a main effect of instructions, $F(1, 28) = 4.60, p < .05, \eta^2_p = .16$; full sample $F(1, 29) = 5.31, p < .05, \eta^2_p = .16$, indicating that logic judgments took almost 1,000 ms longer than belief judgments (4,654 ms vs. 3,781 ms). No other effects in the analysis attained significance.

**Overall analysis.** Table 3 shows accuracy and latency of response averaged across both problem blocks in Experiment 1. A 2 (instructions) × 2 (problem type) × 2 (response type) repeated-measures ANOVA replicated the effect of instructions on accuracy reported in the between-subjects comparison earlier, $F(1, 25) = 5.26, p = .031, \eta^2_p = .17$; full sample, $F(1, 29) = 10.78, p = .003, \eta^2_p = .27$. Once again, participants were more accurate under logical instructions than belief instructions (99% vs. 96%, respectively). There was also a main effect of problem type, $F(1, 25) = 23.20, p < .001, \eta^2_p = .48$, showing greater accuracy with nonconflict than conflict problems (99% vs. 95%, respectively). The interaction between instructions and problem type was not significant in this analysis, $F(1, 25) = 1.73, p = .20, \eta^2_p = .065$, although it did reach significance when the full sample was included in the analysis, $F(1, 29) = 6.24, p = .018, \eta^2_p = .18$.

The analysis of the latency data showed a main effect of instructions, $F(1, 25) = 7.65, p < .05, \eta^2_p = .23$; full sample $F(1, 29) = 5.21, p < .05, \eta^2_p = .15$, replicating the between-participants effect described earlier, whereby logic judgments take longer than belief judgments (4,004 ms vs. 3,252 ms). There was a marginal effect of problem type, $F(1, 25) = 3.74, p = .06, \eta^2_p = .13$; full sample $F(1, 29) = 4.92, p < .05, \eta^2_p = .15$, indicating longer latencies for conflict than for nonconflict problems (3,734 ms vs. 3,522 ms) and a marginal interaction between instructions and problem type, $F(1, 25) = 3.02, p = .09, \eta^2_p = .11$; full sample $F(1, 29) = 3.43, p = .07, \eta^2_p = .11$. Follow-up analyses on this marginal interaction showed that participants took significantly longer to respond to conflict problems compared with nonconflict problems under belief instructions, $F(1, 25) = 9.26, p < .01, \eta^2_p = .27$, but there was no effect of problem type on latencies of response under logic instructions ($F < 1$). This provides additional evidence that, if anything, belief–logic conflict has a greater effect on judgments of conclusion believability than conclusion validity. Finally, there was also a large main effect of response type in this analysis, $F(1, 25) = 19.38, p < .01, \eta^2_p = .44$; full sample $F(1, 29) = 9.04, p < .01, \eta^2_p = .24$, reflecting the tendency for the no responses to take longer than the yes responses (3,780 ms vs. 3,476 ms). None of the remaining interactions in the ANOVA were significant ($F_s < 1$).

The findings from the current study are intriguing. On the one hand, the evidence that belief-based responding takes less time than logical responding is consistent with the DI account, which assumes that beliefs are activated automatically, cueing responses by default. This confirms one of the predictions set out in the
introduction (DP1). However, the other findings of the study are difficult to reconcile with this view and directly contradict the other two predictions that were derived in the introduction (DP2 and DP3). First, we have evidence that belief-based judgments are more rather than less prone to error than logical judgments. Second, the difficulty of belief-based judgments is specific to problems in which there is a conflict with logic, suggesting that logical structure interferes with judgments of belief. In contrast, belief–logic conflict has no influence on judgments of logical validity, a finding that is inconsistent with DP3. This finding is supported by the latency data, which show a trend for an interaction between problem type and instructions, whereby problem type has the greatest impact on belief-based judgments. These findings are more consistent with a DI model in which logic generates the default response. However, it must be noted that the effects reported are small in absolute terms, with overall accuracy on these simple problems under both instructional sets being higher than 95%.

In Experiment 2, we retested the participants from Experiment 1, but this time instead of presenting the problems in blocks with associated instructions, we provided an instructional cue immediately prior to the presentation of each problem. Our aim was to increase the difficulty of the task by mixing belief and logic trials within the same block and determine whether the findings of Experiment 1 could be replicated. Conducting the experiment with the same participants ensured that they were familiar with the instructional manipulation, having already reached a very high level of performance on these problems. Therefore, errors could be clearly attributed to the difficulty of the different instructional sets rather than to any misunderstanding of the instructions. In Experiment 2, we also introduced a new set of conditional arguments in order to eliminate any possibility that the findings in Experiment 1 may be specific to the problem content used.

**Experiment 2**

**Method**

**Design.** Participants received a total of 40 problems, half of which were conflict and half of which were nonconflict problems. Half of the nonconflict problems required a yes answer, and half required a no answer. The conflict items had either valid unbelievable or invalid unbelievable conclusions that required equal numbers of yes/no responses under each instruction type. Each trial involved the presentation of an instructional cue, prior to the presentation of each problem, indicating whether judgments should be based upon beliefs or logic. Half of the problems were preceded by a logic cue, and half of the problems were preceded by a belief cue. This resulted in a 2 (instructions) × 2 (problem type) × 2 (response type) design equivalent to the overall design used in Experiment 1.

**Participants.** The participants were the same as those who took part in Experiment 1. They completed the study during a later testing session.

**Materials and procedure.** A new set of conditional arguments were generated in the same way as Experiment 1. Each conditional was constructed such that the consequent to the factual version was a member of an antonym pair, which could be used to generate a contrary-to-fact version and could also be used as the opposite conclusion to produce problems requiring both yes and no responses (see Table 1).

The participants had previously taken part in Experiment 1 in an earlier testing session during the same day. Therefore, they were reasonably well practiced in applying the different instructions used in Experiment 1 and were given a reminder of the different instructions used in that experiment. They were then presented with the following additional instructions for this part of the study:

Now you will be presented with another series of reasoning problems similar to those in the previous test, but this time your task is to answer according to belief or logic in response to a cue that appears before each problem:

Each problem will be preceded by a “+” to focus your attention, followed briefly by an instructional cue of either the letter B, indicating that you must answer the following problem according to your beliefs, or the letter L, indicating that you must answer according to logic.

Remember, a belief-based response is one that corresponds to what you know to be true in the world, and a logic-based response means that you must assume that each premise is true and respond with the answer that logically follows from the statements presented.

The problem will be presented after the cue, and you must answer either yes by pressing Y or no by pressing N as quickly as possible, just as you did in the first task. Press the space bar when you are ready to start.

The 40 problems were presented in a random order for each participant, half of the trials preceded by a logic cue and half by a belief cue. Each trial consisted of the fixation point, “+” presented for 500 ms, followed by the instructional cue for the trial. The cue was presented for 1,000 ms and was either the letter B, indicating that participants should respond on the basis of beliefs, or L, indicating that they should respond logically. The problem was then presented in the same way as Experiment 1.

**Results and Discussion**

All participants scored at above-chance levels on the conflict problems and were all included in the analysis. Table 4 shows the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Belief instructions</th>
<th>Logic instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response accuracy (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonconflict</td>
<td>78</td>
<td>94</td>
</tr>
<tr>
<td>Conflict</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>Response latency (ms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonconflict</td>
<td>3,523</td>
<td>3,826</td>
</tr>
<tr>
<td>Conflict</td>
<td>3,268</td>
<td>3,857</td>
</tr>
</tbody>
</table>
percentage accuracy rates and latencies for correct responses for each of the problem types for each instructional cue. A 2 (instructions) × 2 (problem type) × 2 (response) repeated-measures ANOVA on the accuracy data revealed a main effect of instructions, F(1, 29) = 20.04, p < .001, ηp² = .41, a main effect of problem type, F(1, 29) = 45.95, p < .001, ηp² = .61, and a significant interaction between these two variables, F(1, 29) = 12.62, p = .001, ηp² = .30. As in Experiment 1, planned comparisons on the interaction revealed a significant difference in accuracy between the conflict and nonconflict problems under belief instructions, F(1, 29) = 43.18, p < .01, ηp² = .60, but just a marginal effect of problem type under logic instructions, F(1, 29) = 4.06, p = .053, ηp² = .12. There was no effect of response type (F < 1), and none of the remaining interactions in the analysis were significant.

We conducted an equivalent analysis on the latencies for correct response. There were two missing data points, which represents less than 1% of the data points used in the analysis. Missing data were replaced by the condition mean. As in Experiment 1, there was a main effect of instructions, F(1, 29) = 7.00, p < .05, ηp² = .19, indicating that participants took longer to respond under logic instructions than under belief instructions (3,841 ms vs. 3,395 ms). There was also a main effect of response type, F(1, 29) = 12.34, p < .01, ηp² = .29, with participants taking significantly longer to respond no than yes (3,773 ms vs. 3,464 ms). There was no effect of problem type (F < 1), and none of the remaining interactions were significant.

In Experiment 2, some of the key findings from Experiment 1 were replicated. In terms of accuracy, participants made more errors when instructed to respond on the basis of beliefs, and this effect was magnified when there was a conflict between the logical and belief-based responses. Once again, this appears to show that judgments concerning the believability of a conclusion are influenced by the logic of the conclusion, but judgments about the validity of a conclusion are relatively uninfluenced by beliefs on these simple arguments. Of note in Experiment 2 is that the use of mixed instructional cues presented immediately before each argument reduced accuracy compared with the blocked presentation used in the previous study. It had its greatest impact on conflict problems when participants were under belief instructions. This suggests that participants are more sensitive to manipulations that increase task demands when they are responding to problems on the basis of beliefs, whereas participants are relatively unaffected when responding logically.

The accuracy data are, if anything, consistent with an account in which logical processing is primary and must be inhibited in favor of judgments based upon background knowledge. However, under this view, one would expect some evidence from the latency data to indicate that reasoning under belief instructions takes more time. In fact, the findings from Experiments 1 and 2 show the opposite: logical reasoning consistently takes longer than belief-based reasoning. How might these two sets of findings be reconciled? One possibility is that belief instructions encourage a strategic approach to the task for some participants. Consider, for example, the following problem from Experiment 2:

If you pick up a feather, then it will be heavy.
Suppose you pick up a feather.
Will it be heavy?

Under logical instructions, participants have to process the conditional premise and the categorical premise and combine them in order to judge whether the conclusion logically follows or not. However, under belief instructions, participants need only to inspect the categorical premise (“you pick up a feather”), the conclusion (“the feather is heavy”), and the beliefs about whether feathers are heavy in order to judge that the conclusion is unbelievable. Given that participants were cued prior to problem presentation, it is quite possible that under belief instructions some participants make use of such a strategic shortcut to the response, resulting in quicker processing times. Our aim in Experiment 3 was to compare the time course of belief-based and logical processing on these arguments, while minimizing the participants’ opportunity to develop a strategic shortcut to the response.

**Experiment 3**

In Experiment 3, the instructional cue was provided after, rather than prior to, problem presentation. In this design, participants were unable to focus only on specific problem components after being presented with a belief instruction cue.

**Method**

**Design.** The design of the experiment was similar to Experiments 1 and 2. Participants received a total of 48 MP arguments. Type of instruction (logic or belief), problem type (conflict and no conflict), and nature of response (yes or no) were manipulated within subjects.

**Participants.** The participants were 28 women and four men who were undergraduate students at the University of Plymouth. The volunteer participants took part in the study in return for course credit that was part of a course requirement.

**Materials and procedure.** Forty-eight different conditional arguments were constructed. The content and the conclusions to each argument were modified in order to construct four variations of the same content, corresponding to the four categories of problems shown in Table 1. This resulted in a total of 192 different problems or four sets of 48 problems. Each set of 48 problems consisted of conditionals with different content, 24 conflict problems and 24 nonconflict problems. Half of these problems were presented with a belief instruction and half with a logic instruction. Under each instructional set, half of the problems required a yes response and half required a no response. This represented six problems in each cell of the design. Participants received one of the four different sets of problems, with eight participants receiving each set.

Participants were instructed in a similar way to participants in Experiment 1. They were informed that they would be presented with 48 problems and that they would be required to respond on the basis of whether a conclusion was either believable or logically followed. The belief instructions emphasized that participants should respond according to the believability of the conclusion. The instruction was followed by two examples, one conflict problem and one nonconflict problem illustrating the correct response under this instructional set. The logic instructions provided a definition of logical validity, followed by the same two examples together with details of the correct logical response.

Each trial consisted of the presentation of the conditional premise alone. Participants then pressed the space bar on a key-
board; the conditional premise disappeared and was replaced by the categorical premise, the conclusion, and the response options, which acted as the instructional cue. Presenting the problems in this way ensured that participants could engage in limited logical processing until the space bar was pressed and the categorical premise, conclusion, and instructional cue were available. An example of problem presentation follows:

If an animal is a bird, then it can fly.
The animal is a bird; therefore, it can fly.

**BELIEVABLE (VALID) → UNBELIEVABLE (INVALID)**

The response options were either believable/unbelievable or valid/invalid. Participants were instructed that the response options shown at the bottom of the screen indicated whether they should respond on the basis of the validity of the conclusion or the believability of the conclusion. After the instructions were presented, participants were shown eight practice arguments: four conflict and four nonconflict problems, half of which were accompanied by validity response options and half by belief response options. Feedback provided after each practice trial indicated whether their response was correct or incorrect. After completion of the practice trials, participants moved on to the main part of the experiment by pressing the space bar.

**Results and Discussion**

Three participants were eliminated from the analysis because they scored below chance on the conflict problems. Consequently, the main analyses reported include the remaining 29 participants, although all significant effects are reported for the full sample also. A 2 (instruction) × 2 (problem type) × 2 (response type) ANOVA on response accuracy revealed a marginal main effect of instructions, $F(1, 28) = 3.27, p = .08, \eta_p^2 = .11$; full sample $F(1, 31) = 4.27, p = .047, \eta_p^2 = .12$, which, as Table 5 shows, replicates the finding of Experiments 1 and 2 that more errors are made under belief than under logic instructions (84% vs. 92% accuracy). There was also a main effect of problem type, $F(1, 28) = 24.99, p < .001, \eta_p^2 = .48$; full sample $F(1, 31) = 32.00, p < .001, \eta_p^2 = .51$, reflecting greater accuracy on nonconflict than conflict problems (96% vs. 79%). Once again, these variables showed a significant interaction, $F(1, 28) = 6.97, p = .013, \eta_p^2 = .20$; full sample $F(1, 31) = 8.07, p = .008, \eta_p^2 = .21$. Table 5 shows that the interaction revealed itself as a greater effect of problem type under belief instructions, $F(1, 28) = 11.5, p < .01, \eta_p^2 = .33$, than under logic instructions, $F(1, 28) = 3.54, p = .07, \eta_p^2 = .11$, as demonstrated by planned comparisons on the interaction. This once again indicates that judgments about conclusion believability are more influenced by belief and logic conflict than by judgments concerning conclusion validity. None of the remaining main effects or interactions were significant.

Table 5 also shows latency of response under each instructional condition. In order to minimize the number of missing cells, we collapsed over response type in the analysis of latencies for correct responses. There were two missing data points, and these were replaced by the condition mean. A 2 (instructions) × 2 (problem type) ANOVA revealed a main effect of problem type, $F(1, 28) = 8.04, p < .01, \eta_p^2 = .22$; full sample $F(1, 31) = 7.2, p < .05, \eta_p^2 = .19$, and a main effect of instructions, $F(1, 28) = 9.52, p < .01, \eta_p^2 = .25$; full sample $F(1, 31) = 10.8, p < .01, \eta_p^2 = .26$. Unlike Experiments 1 and 2, this experiment showed that participants took significantly longer to respond under belief instructions than logic instructions (3,368 ms vs. 3,076 ms). Table 5 shows that the difference in latencies between the conflict and nonconflict problems appears to be greater under belief than logic instructions. Although the interaction between problem type and instructions was not significant, $F(1, 28) = 2.45, p = .13, \eta_p^2 = .08$, a simple effects analysis on the interaction revealed an effect of problem type under belief instructions, $F(1, 28) = 7.99, p < .01, \eta_p^2 = .22$, but no effect under logic instructions ($F < 1$).

Our aim in Experiment 3 was to determine whether providing the instructional cue following problem presentation would eliminate the latency effects observed in Experiments 1 and 2. Here, where the cue was provided before the problem, belief judgments were made more quickly than logic judgments. We suggested that this difference may arise not because belief-based processing is inherently quicker than logic-based processing but because the preceding cue may encourage participants to use a shortcut strategy that does not involve processing the full conditional argument. In this study, where such a shortcut strategy could not be applied, belief-based judgments took substantially longer than judgments about conclusion validity. Experiment 3 has also replicated many of the findings of Experiments 1 and 2; participants made more errors under belief instructions, and these errors primarily were confined to problems in which logic and belief were in conflict. In contrast, the effect of conflict on logical judgments was substantially smaller. This finding is mirrored in the latency data, which indicate that participants take significantly longer to process conflict problems than nonconflict problems under belief instructions, but this difference is absent when participants receive logic instructions.

**Experiment 4**

The findings reported thus far are inconsistent with the DI dual-process account. In many respects, this is surprising, given the number of findings in the literature that seem to support the view that belief-based responding is rapid and automatic and that it is the default response. De Neys (2006), for example, has shown that belief bias in syllogistic reasoning increases when participants are put under memory load. Similar effects have been reported with speeded task manipulations (Evans & Curtis-Holmes, 2005). One of the characteristics of the syllogistic reasoning tasks used in these studies is that conclusion believability is manipulated directly; that is, the believability of the conclusion is manipulated rather than the believability of the premises. In the conflict prob-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Belief instructions</th>
<th>Logic instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conflict</td>
<td>Nonconflict</td>
</tr>
<tr>
<td>Response accuracy (%)</td>
<td>71</td>
<td>97</td>
</tr>
<tr>
<td>Response latency (ms)</td>
<td>3,558</td>
<td>3,128</td>
</tr>
</tbody>
</table>
lems we have used here, the believability of the conclusion is a
function of the participants’ belief in the conditional premise. 
Some recent evidence has shown that this might be an important 
task difference that explains the absence of any effect of speeded 
task on belief bias in conditional reasoning (Evans, Handley, & 
Bacon, 2009). Our primary aim in Experiment 4 was to examine 
whether our findings extended to problems in which conclusion 
believability was independent of belief in the major premises.

In Experiment 4, a new set of problems was constructed in 
which nonsense linking terms were used in order to ensure that 
participants had no pre-existing beliefs concerning the believability 
of the conditional premise. Table 6 gives an example of one of 
the problem contents employed. In addition, problems of the kind 
used in Experiment 3 were administered.

**Method**

**Design.** The design of the experiment was similar to that of 
Experiment 3. Participants received a total of 48 MP arguments.

Type of instruction (logic or belief), problem content (real or 
nonsense middle term), problem type (conflict and nonconflict),
and nature of response (yes or no) were manipulated within par-
ticipants.

**Participants.** The participants were 28 women and four men 
who were undergraduate students at the University of Plymouth. 
The volunteer participants took part in the study in return for 
course credit that was part of a course requirement.

**Materials and procedure.** As in Experiment 3, 48 condi-
tional arguments were constructed, each argument with different 
content. Half of these MP arguments contained nonsense linking 
terms, and half of the arguments contained real linking terms (of 
the kind used in Experiment 3). Four versions of each of the 48 
arguments were constructed as illustrated in Table 6. This resulted 
in four sets of 48 arguments, each consisting of 24 problems with 
real linking terms and 24 with nonsense linking terms. Half of the 
problems were presented with an accompanying belief instruc-
tional cue and half of the problems with a logic instructional cue.

Half of the problems were conflict, and half were nonconflict, and 
of these there were equal numbers requiring yes and no responses,
resulting in three problems within each cell of the design. The 
problem instructions, problem presentation, and the experimental 
procedure were identical to those in Experiment 3.

**Results and Discussion**

Percentage accuracy and latencies for correct responses are 
shown in Table 7. Performance in Experiment 4 was generally 
poorer than in Experiment 3, with higher overall error rates and 
longer latencies on conflict items. Seven participants scored below 
chance on the conflict items, a higher proportion than in any of 
the other studies reported, and these participants were eliminated 
from the main analysis. One possible reason for the increased difficulty 
of the problems in Experiment 4 is that different problem 
structures were employed, so participants needed to identify and 
differentiate between the types of problems presented on each trial.
This additional task demand may have contributed to increased 
latenities and higher error rates. A 2 (instruction) × 2 (type of 
linking term) × 2 (problem type) × 2 (response repeated-
measures ANOVA) on problem accuracy revealed a main effect of 
instructions, *F*(1, 24) = 5.03, *p* = .011, *η*₂ = .21; full sample (*F*
(1, 31) = 5.31, *p* = .028, *η*₂ = .13, reflecting greater accuracy under 
logical than belief instructions (88% vs. 76%). There was also a 
main effect of problem type, *F*(1, 24) = 92.30, *p* < .001, *η*₂ = .78; 
full sample (*F*(1, 31) = 107.01, *p* < .001, *η*₂ = .75, reflecting 
greater accuracy on nonconflict than on conflict problems (94% vs. 
69%). The interaction between instructions and problem type was 
marginally significant, *F*(1, 24) = 3.90, *p* = .060, *η*₂ = .14; full 
sample (*F*(1, 31) = 3.38, *p* = .076, *η*₂ = .10, but would be 
comfortably so on a one-tailed test. As in the earlier experiments, 
this interaction reflects a larger effect of problem type under belief 
instructions, *F*(1, 24) = 39.87, *p* < .01, *η*₂ = .62, than under logic 
instructions, *F*(1, 24) = 15.51, *p* < .01, *η*₂ = .39. There was no 
effect of problem content on judgment accuracy, *F*(1, 24) = 2.59, 
and none of the remaining main effects or interactions were 
significant.

As in Experiment 3, in order to minimize the number of missing 
cells, we collapsed the response type variable in the analysis of 
latenities. This left seven missing data points, representing 3.5% of 
the total number of data points in the analysis, which were replaced 
by the mean for the relevant condition. The resulting 2 (instruc-
tion) × 2 (real vs. nonsense linking term) × 2 (problem type) 
ANOVA replicated the main effect of instruction shown in Experi-
ment 3, *F*(1, 24) = 12.51, *p* < .01, *η*₂ = .34; full sample (*F*(1, 
31) = 13.03, *p* < .01, *η*₂ = .29. As Table 7 shows, belief-based 
judgments took longer than judgments of conclusion validity 
(4,459 ms vs. 3,644 ms). There was a main effect of content, *F*(1, 
24) = 15.27, *p* < .01, *η*₂ = .39; full sample (*F*(1, 31) = 17.89, *p* < .01, 
*η*₂ = .37, reflecting longer latencies on the problems with 
nonsense linking terms than on those with the realistic linking 
terms (4,459 ms vs. 3,644 ms). There was also a main effect of 
problem type, *F*(1, 24) = 32.75, *p* < .01, *η*₂ = .57; full sample 
(*F*(1, 31) = 40.01, *p* < .01, *η*₂ = .56, and an interaction between 
instructions and problem type, *F*(1, 24) = 17.58, *p* < .01, *η*₂ = .42; 
full sample (*F*(1, 31) = 12.68, *p* < .01, *η*₂ = .29. Once again, 
this reflected a larger effect of problem type under belief instruc-
tions, *F*(1, 24) = 42.49, *p* < .01, *η*₂ = .64, than under logic 
instructions, *F*(1, 24) = 3.62, *p* = .07, *η*₂ = .13. None of the 
remaining interactions were significant (*F* < 1).

---

**Table 6**

*Example of the Different Types of Problems Used in Experiment 4 With the Responses Required Under Logic or Belief-Based Instructional Sets*

<table>
<thead>
<tr>
<th>Conflict problem</th>
<th>Nonconflict problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument A</td>
<td>Argument B</td>
</tr>
<tr>
<td>All zabs can walk.</td>
<td>All zabs can swim.</td>
</tr>
<tr>
<td>Whales are zabs.</td>
<td>Whales are zabs.</td>
</tr>
<tr>
<td>Therefore, whales can walk.</td>
<td>Therefore, whales can swim.</td>
</tr>
<tr>
<td>Response Logic: YES</td>
<td>Logic: YES</td>
</tr>
<tr>
<td>Belief: NO</td>
<td>Belief: YES</td>
</tr>
<tr>
<td>Argument C</td>
<td>Argument D</td>
</tr>
<tr>
<td>All zabs can walk.</td>
<td>All zabs can swim.</td>
</tr>
<tr>
<td>Whales are zabs.</td>
<td>Whales are zabs.</td>
</tr>
<tr>
<td>Therefore, whales can swim.</td>
<td>Therefore, whales can walk.</td>
</tr>
<tr>
<td>Response Logic: NO</td>
<td>Logic: NO</td>
</tr>
<tr>
<td>Belief: YES</td>
<td>Belief: NO</td>
</tr>
</tbody>
</table>
Experiment 4 was designed to examine whether the effect of conflict under belief instructions extended to problems in which believability judgments depended solely upon the believability of the conclusion. To this end, we devised a set of problems in which nonsense middle terms were contained; these types of problems are more consistent with the sorts of problems investigated thus far in the dual-process literature. In Experiment 4, all of the key findings of Experiment 3 were replicated, and none of these effects interacted with problem content. Once again, judgments concerning conclusion believability took longer than judgments about conclusion validity, and the presence of a conflict between logic and belief had a significant effect on latencies for belief judgments but no effect on latencies for validity judgments. These findings were mirrored in the accuracy data, which showed a much larger effect of conflict under belief instructions.

The absence of any modifying impact of content on these effects is informative. What it shows, perhaps surprisingly, is that even on a task in which judgments of believability can be made with reference only to the content of the conclusion, the logical structure of the argument continues to interfere with judgments. This suggests that at least on these simple problems, the mechanism involved in processing the underlying logical structure of the arguments pre-empts belief-based reasoning.

Up to now, we have focused on simple MP arguments. Often, logical performance on this inference form is close to ceiling, and it has been argued that MP is supported by direct inference rules encoded as part of the lexical entry for “if” (see, for example, Braine & O’Brien, 1991). Perhaps certain inferences, such as MP, are genuinely default and consequently activated prior to the activation of relevant knowledge, which causes interference with belief-based judgments but no interference in the reverse direction, as observed. Our aim in Experiment 5 was to examine the impact of belief–logic conflict on more complex disjunctive argument forms.

Experiment 5

In Experiment 5 we employed arguments of the following kind:

Either the sky is blue, or it is green,
The sky is not blue.
Therefore, the sky is green.

Disjunctive arguments of this kind are referred to as denial inferences because the categorical premise denies one of the constituents of the major premise. We also included affirmation inferences, which involve reasoning from the affirmation of one constituent (e.g., “the sky is green”) to the denial of the other component (“therefore, the sky is not blue”). Both types of disjunctive argument have been shown to result in more errors than the MP inference (see, for example, Evans, Newstead, & Byrne, 1993). Table 8 gives an example of the conflict and nonconflict items associated with each argument form. As Table 8 shows, the use of both argument types ensures that the polarity of the conclusion presented is not confounded with logical validity. We predicted that the key finding of an interaction between instruction and problem type would extend to disjunctive arguments.

Method

Design. Participants received a total of 64 MP conditionals and 64 disjunctive syllogisms. As in previous studies, instruction type (logic or belief), problem type (conflict or nonconflict), and nature of response (yes or no response required) was manipulated. In addition, we also manipulated argument type (MP vs. disjunctive syllogism). All manipulations were within participants.

Participants. The participants were 30 female and five male undergraduate students from the University of Plymouth who participated in exchange for course credit.

Materials and procedure. Participants received a total of 128 problems: 64 disjunctive arguments and 64 MP arguments. Half of the problems were conflict problems, and half were nonconflict problems.

Instructions were similar to the ones used in Experiment 3. Participants were informed that they would be presented with 128 reasoning problems to which they should respond according to the believability or validity of the conclusion. The instructions were presented before the experiment both on paper and on a computer screen. Examples of both MP and disjunctive syllogisms were given.

As in Experiment 3, the conditional premise was displayed first. The participant then pressed the space bar, and the conditional premise disappeared and was replaced by the categorical premise, the conclusion, and the response options. As in Experiment 3, the response options acted as the instructional cue; that is, when a logic-based response was required, the response options were VALID and INVALID. If on the other hand a belief-based response was required, the screen would show BELIEVABLE and UNBELIEVABLE. Participants received 16 practice trials with feedback consisting of examples of all the key problem types. If a participant responded incorrectly to a practice item, a message appeared stressing that the participant should ask the experimenter to explain the instructions again. Upon completion of the 16 practice trials, the participant initiated the main part of the experiment by pressing the space bar.

Results and Discussion

Six participants were removed from the analysis because they scored below chance on conflict items. All analyses were performed on the remaining 29 participants; however, in line with previous studies, we also report the statistical analysis for the full

<table>
<thead>
<tr>
<th>Variable</th>
<th>Belief instructions</th>
<th>Logic instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conflict</td>
<td>Nonconflict</td>
</tr>
<tr>
<td>Response accuracy (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real middle term</td>
<td>59</td>
<td>93</td>
</tr>
<tr>
<td>Nonsense middle term</td>
<td>58</td>
<td>91</td>
</tr>
<tr>
<td>Response latency (ms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real middle term</td>
<td>4,622</td>
<td>3,311</td>
</tr>
<tr>
<td>Nonsense middle term</td>
<td>5,564</td>
<td>3,890</td>
</tr>
</tbody>
</table>

Table 7

Accuracy and Latency of Response Under Belief and Logic Instructions for Each Content and Problem Type in Experiment 4 (N = 25)
sample. We analyzed the accuracy data using a 2 (problem type: conflict vs. no conflict) × 2 (instruction: belief vs. logic) repeated-measures ANOVA. We found a main effect of problem type, $F(1, 28) = 18.95, p < .001, \eta^2_p = .40$; full sample $F(1, 34) = 30.79, p < .001, \eta^2_p = .48$, showing that accuracy was lower for conflict items than for nonconflict items (84% vs. 93%). A main effect for argument type was also found, $F(1, 28) = 18.95, p < .001, \eta^2_p = .40$; full sample $F(1, 34) = 22.71, p < .001, \eta^2_p = .40$, confirming that performance was higher for MP arguments than for disjunctive syllogisms (91% vs. 86%). As in the previous studies, we found that problem type and instructions interacted, $F(1, 28) = 20.99, p < .001, \eta^2_p = .43$; full sample $F(1, 34) = 16.82, p < .001, \eta^2_p = .33$, showing that under belief instructions, there was a bigger performance drop for conflict items ($M_{\text{conflict}} = 80$% vs. $M_{\text{nonconflict}} = 94$%) than under logic instructions ($M_{\text{conflict}} = 89$% vs. $M_{\text{nonconflict}} = 91$%).

In order to verify that the Problem Type × Instructions interaction was present for both argument types, we conducted a 2 (problem type) × 2 (instruction) repeated-measures ANOVA for MP arguments and disjunctive syllogisms separately. The analysis for MP arguments confirmed the Problem Type × Instructions interaction was significant, $F(1, 28) = 4.92, p = .035, \eta^2_p = .15$; full sample $F(1, 34) = 6.76, p = .014, \eta^2_p = .17$. The separate analysis for disjunctive syllogisms also resulted in a significant Problem Type × Instructions interaction, $F(1, 28) = 14.39, p < .001, \eta^2_p = .34$; full sample $F(1, 34) = 15.78, p < .001, \eta^2_p = .32$. Cell means by argument type can be found in Table 9.

A 2 (problem type) × 2 (instruction) × 2 (argument type) repeated-measures ANOVA on the latencies for correct responses revealed a main effect of problem type, $F(1, 28) = 15.09, p = .001, \eta^2_p = .35$; full sample $F(1, 34) = 23.35, p < .001, \eta^2_p = .41$, indicating that people took longer to respond to conflict items than to nonconflict items (4,424 ms vs. 4,152 ms). There was also a main effect of instruction, $F(1, 28) = 23.10, p < .001, \eta^2_p = .45$; full sample $F(1, 34) = 10.95, p = .002, \eta^2_p = .24$, which indicated that responding under a belief-based instructions took longer than responding under logic instructions (4,454 ms vs. 4,122 ms). A main effect of argument type was also found, $F(1, 28) = 127.4$, 

Table 8

<table>
<thead>
<tr>
<th>Problem Type A</th>
<th>Nonconflict problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument A</td>
<td></td>
</tr>
<tr>
<td>Either the sky is blue, or it is green.</td>
<td></td>
</tr>
<tr>
<td>Suppose the sky is blue.</td>
<td></td>
</tr>
<tr>
<td>Does it follow that the sky is green?</td>
<td></td>
</tr>
<tr>
<td>Response Logic: YES</td>
<td></td>
</tr>
<tr>
<td>Belief: NO</td>
<td></td>
</tr>
<tr>
<td>Argument C</td>
<td></td>
</tr>
<tr>
<td>Either the sky is blue, or it is green.</td>
<td></td>
</tr>
<tr>
<td>Suppose the sky is blue.</td>
<td></td>
</tr>
<tr>
<td>Does it follow that the sky is not green?</td>
<td></td>
</tr>
<tr>
<td>Response Logic: NO</td>
<td></td>
</tr>
<tr>
<td>Belief: YES</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem Type B</th>
<th>Conflicts problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument A</td>
<td></td>
</tr>
<tr>
<td>Either the sky is blue, or it is green.</td>
<td></td>
</tr>
<tr>
<td>Suppose the sky is green.</td>
<td></td>
</tr>
<tr>
<td>Does it follow that the sky is not blue?</td>
<td></td>
</tr>
<tr>
<td>Response Logic: YES</td>
<td></td>
</tr>
<tr>
<td>Belief: NO</td>
<td></td>
</tr>
<tr>
<td>Argument C</td>
<td></td>
</tr>
<tr>
<td>Either the sky is blue, or it is green.</td>
<td></td>
</tr>
<tr>
<td>Suppose the sky is green.</td>
<td></td>
</tr>
<tr>
<td>Does it follow that the sky is blue?</td>
<td></td>
</tr>
<tr>
<td>Response Logic: NO</td>
<td></td>
</tr>
<tr>
<td>Belief: YES</td>
<td></td>
</tr>
</tbody>
</table>

Table 9

<table>
<thead>
<tr>
<th>Variable</th>
<th>Belief instructions</th>
<th>Logic instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modus ponens</td>
<td>Conflict</td>
<td>Nonconflict</td>
</tr>
<tr>
<td>Response accuracy (%)</td>
<td>86</td>
<td>96</td>
</tr>
<tr>
<td>Response latency (ms)</td>
<td>3,908</td>
<td>3,438</td>
</tr>
<tr>
<td>Disjunctive syllogisms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response accuracy (%)</td>
<td>75</td>
<td>93</td>
</tr>
<tr>
<td>Response latency (ms)</td>
<td>5,500</td>
<td>4,972</td>
</tr>
</tbody>
</table>

HANDLEY, NEWSTEAD, AND TRIPPAS
of shortcut strategies on these problems; that is, participants could generate a response without processing the major premise. In Experiments 3–5, in which the response cue was presented following problem presentation, belief judgments took longer than logic judgments. As in the analysis of errors, the increased latency of responding under belief instructions was limited to problems in which belief and logic were in conflict. Once again this suggests that logical processing on these problems is primary and pre-empts belief-based processing.

Before considering the theoretical implications of these findings, we consider whether there could be a simple methodological explanation for the results. Is it possible, for example, that some participants misunderstood what they were instructed to do and confused the instruction to respond according to their beliefs with the instruction to respond logically? Perhaps they understood the instruction as a request to respond according to their beliefs about the logical validity of the conclusion? This would of course result in a higher error rate under belief instructions on problems in which the belief and logic responses differed. There are a number of reasons why we do not think this explanation could account for our findings. First, we used a strict criterion of eliminating participants who scored at the level of chance on the conflict problems, consequently removing those individuals who may have misconstrued the instructions. Second, the difference in performance on conflict problems was present even at high levels of accuracy. In Experiment 1, for example, in which problems were presented in blocks, accuracy was above 90% in every condition. Third, participants were instructed very clearly regarding the type of response they should generate under each instructional condition. The belief and logic instructions were presented contiguously, with the differing response requirements for belief and logic trials clearly contrasted and practice problems with feedback provided in Experiments 3–5. Given such a careful approach to instructing participants, we would be surprised if there was a misunderstanding of the task requirements. Finally, the latency analyses, which broadly support the analysis of response accuracy, were based upon latencies for correct responses, where participants are responding appropriately to task instructions. Taken together, it seems unlikely that the findings can be fully explained as arising from a misconstrual of the task requirements under each instructional condition.

The experiments predominantly employed within-participants designs. Is it possible that participants were confused from trial to trial concerning the type of response required? It is, after all, quite unusual to require participants to ignore the structural characteristics of a problem in favor of a response based upon general knowledge. We know that the requirement to switch between different tasks is associated with significant costs (see, for example, Monsell, 2003), and there is some evidence to suggest that performance on certain tasks may be more susceptible to task-switching effects than performance on other tasks. For example, switch costs associated with automatic processes (such as word naming on the Stroop task) are greater than switch costs associated with more controlled processing (such as color naming on the Stroop; see Wylie & Allport, 2000). These costs have been shown to extend to subsequent word-naming nonswitch trials. This finding has been explained by the claim that the automatic processes associated with word reading must be inhibited during color-naming trials. This inhibition has a knock-on effect on subsequent
word-naming trials, leading to long-lasting switch costs. In contrast, because word naming does not normally require the inhibition of color naming, there is less persisting inhibition of color naming left over from the word-reading trials. Consequently, there are asymmetric switch costs associated with trials that are more reliant on automatic processing. Is it possible that belief-based reasoning is, in fact, an automatic process, and the longer latencies and increased errors on these trials are carryover effects arising from long-lasting inhibitory processes engaged during logical reasoning trials? There are two reasons why we think that it is not a satisfactory explanation of our findings. Although there is a significant accuracy cost associated with mixing trials compared with a blocked design (see Experiment 1 vs. Experiment 2), the key findings also emerged when problems were presented within a single block. When performance on the first block of problems was examined in Experiment 1, the difference between instructional conditions remained, as did the interaction between problem type and instructional set. This shows that the effect is not exclusive to within-participants designs, and consequently any explanation based upon asymmetric switch costs cannot fully account for our data. Second, recent evidence suggests that where inhibition of beliefs occurs during logical reasoning it extends only to beliefs specifically related to the particular problem content used. There is no evidence that memory access is generally impaired after solving conflict problems (see De Neys & Franssens, 2009). This finding suggests that inhibitory processes, in the case of reasoning tasks, operate on very specific beliefs, and consequently generalized carryover effects from logic to belief trials are unlikely to explain the findings here.

Perhaps the repeated presentation of similar argument structures encouraged the development of automaticity in the application of an inference rule that would otherwise be applied with much more effort in isolation? The evidence from text comprehension studies suggests that inferences can be drawn automatically in the context of individual presentation, suggesting that repetition is not a prerequisite for the development of automaticity. In addition, we would argue that our design was such that repeated presentation would not naturally result in the development of automatic responding over time. As discussed earlier, the effects extended to a within-participants design, in which the interspersing of different problem structures and instructional requirements would reduce repetition. In addition, the effects were robust in Experiment 5, where the presentation of three distinct logical structures would have reduced repetition considerably.

Finally we consider whether the effects could be attributed to the content employed in the experimental problems. Could it be that participants were simply uncertain about the believability of the conclusions presented to them, which consequently resulted in their making more errors in the belief condition? There are two reasons why we think that this is not a satisfactory explanation of the findings. First, the belief–logic problems were constructed around contrary-to-fact premises, in which the premises are empirically false and consequently do not vary in terms of graded belief. Therefore, it is unlikely that there would be any uncertainty about the believability of presented conclusions. In addition, we presented all the conclusions from both factual and contrary-to-fact premises, ensuring that the same conclusions were evaluated on both conflict and nonconflict problems. If participants were uncertain about the believability of the conclusions under belief instructions, then this uncertainty should have extended to both factual and contrary-to-fact arguments. Across all four studies, there was no difference between conditions on nonconflict problems. This shows that it cannot be differences in underlying strength of beliefs driving the effect.

The evidence shows quite clearly that conclusion validity interferes with judgments of belief, but conclusion believability has a limited degree of interference with judgments of validity. This suggests that certain logical inferences are accomplished relatively automatically and are immune to the influence of beliefs. In contrast, judging conclusion believability appears to be a slower process that is subject to interference from logical analysis. This pattern is exactly the opposite of what would be expected if belief-based judgments were based on fast, automatic, Type 1 processes and logic-based judgments were based on slow, deliberate, Type 2 processes.

How might one reconcile these findings with a DI account of belief bias? After all, substantial evidence suggests that reasoning independently of beliefs is a cognitively challenging task that draws on cognitive capacity and inhibitory resources. It is important, however, to note that in all cases, such evidence is based upon study designs in which participants are given strict logical instructions that emphasize the importance of the participants assuming the premises and only drawing necessary conclusions. Recently Stanovich and West (2008) have made an important distinction in the dual process literature between algorithmic levels and intentional levels of processing. The algorithmic level is concerned with the information processing procedures required to carry out a given task; the efficiency of these procedures being constrained by the processing capacity of the system. The intentional level reflects the goals of the organism in terms of what it is trying to achieve relative to its knowledge structures and beliefs. Intentional level goals may be manipulated through structural cues or by instructions that encourage decontextualized thinking (Stanovich & West, 2008). Under this view, Type 2 processing is only required when the task is constrained at the intentional level through instructional cues, and it is only in these circumstances that effortful processing is engaged. The relationship between measures of cognitive capacity and reasoning performance vanishes when such cues are not provided (see, for example, Evans et al., 2010). This indicates that successful application of a set of instructions may be the important factor when one is determining the degree of effort required on a task, rather than the specific requirement to reason independently of beliefs.

The novelty of the studies presented here rests on the use of both logic and belief-based instructions, the latter being used for the first time in the study of belief-based reasoning. It is quite possible that the evidence cited as supporting the role of Type 2 processes in participants’ ability to resist belief bias reflects the demanding nature of applying a set of specific instructions, which must be maintained in mind during the course of problem solving. Of course, if this conjecture is right, then we would expect accuracy on belief-based judgments to be related to measures of cognitive capacity and equally affected by secondary loads or speeded tasks as logical judgments, when explicit instructions are presented.

---

1 The authors would like to thank Christoph Klauser for suggesting this alternative explanation for our findings.
These intriguing hypotheses remain to be tested but suggest a number of avenues for future research.

The evidence of greater difficulty for belief-based reasoning compared with logical reasoning is inconsistent with one of the key assumptions of DI accounts: the proposal that beliefs influence reasoning in an automatic way through the default activation of associations in long-term memory. While undoubtedly there are many beliefs that may come to mind fairly rapidly, recent evidence suggests that even very basic beliefs are not activated by default when an individual is processing information in context. For example, recent research has shown that people do not automatically ascribe beliefs to agents when observing another person’s behavior (Apperly, Riggs, Simpson, Chiavarino & Samson, 2006) and that thinking about other people’s beliefs is often an effortful and demanding process (Apperly, Samson & Humphreys, 2005; McKinnon & Moscovitch, 2007). In the domain of reasoning, performance tasks whose solution depends heavily on context appear to draw upon domain general resources. For example, performance on deontic versions of the selection task is impaired under conditions of cognitive load (McKinnon & Moscovitch, 2007) and is related to individual differences in general intelligence (Newstead et al., 2004). In addition, recent research has shown that among preadolescent children, biased responding on reasoning tasks is more, rather than less, common in participants who score higher on measures of fluid intelligence (Morsanyi & Handley, 2008). These findings suggest that reasoning about beliefs is often an effortful process, relying upon the integration of relevant knowledge with aspects of problem structure in working memory in order to generate a novel response. The demanding nature of this process is further supported by recent work that shows that autistic children are less likely to integrate relevant knowledge in reasoning about conditional relations (McKenzie, Evans, & Handley, 2010). This finding is in line with recent claims that autism is characterized by a complex processing deficit that impacts primarily on the capacity to integrate information in working memory (Minshew, Goldstein & Siegel, 1997).

This analysis suggests that evaluating conclusions with respect to beliefs is a Type 2 process rather than a Type 1 process. However, a question remains concerning the status of the logical processing involved in the problems presented here. It is important to recognize that in most of the studies, the arguments employed were based upon MP, an inference that is readily drawn from both abstract and realistic conditional premises. Perhaps this inference is unique, being drawn automatically in the process of comprehending the premises; it is a default inference that must be inhibited for a belief-based judgment to be made. This claim is consistent with the evidence that MP is drawn spontaneously in text processing (Rader & Sloutsky, 2002) and the view that MP is a direct rule of inference that is automatically activated when its form is matched by the structure of a problem (Braine & O’Brien, 1991). Although it may be tempting to assume that MP has a special default inference status, the evidence from Experiment 5 shows clearly that the effects extend to more complex disjunctive arguments. While certain disjunctive inferences have also been shown to be drawn in the process of text comprehension (Lea et al., 1990), it is worth noting that reasoning from disjunctive arguments took substantially longer and resulted in significantly more errors than reasoning from conditional premises. This demonstrates a substantial increase in problem difficulty in line with findings reported elsewhere (see Evans, Newstead & Byrne, 1993). In addition, there is good experimental and neuroimaging evidence that suggests substantial overlap in the processes underlying reasoning from simple argument forms such as MP and more complex argument forms such as modus tollens (MT). For example, MP and MT are equally affected by the presence or availability of disabling conditions, a pattern that one would not expect if MP really were a unique default inference (see, for example, Byrne, 1989). Recent neuroimaging work also suggests substantial overlap in areas of neural activation for these argument forms (Noveck, Goel, & Smith, 2004). Common activation in left parietal frontal networks is also associated with relational reasoning and reasoning from more complex syllogistic reasoning arguments (Goel & Dolan, 2003). This shows that simple inferences such as MP involve the same underlying brain areas as more complex reasoning problems, a finding that is not consistent with the idea that simple inferences are accomplished through a distinct set of processes. In addition, variability in MP rates of inference is often associated with measures of cognitive ability, a finding that is difficult to reconcile with the view that MP is uniquely an automatically activated default inference (Newstead et al., 2004).

In our view, accommodating these findings requires an alternative model to the traditional DI account with its emphasis on default beliefs and effortful logical processing. We favor an alternative account in which the processes underlying different responses are assumed to operate in competition with one another. The process that completes first cues a response, which, depending upon task instructions, available resources, and time, may need to be inhibited in favor of an alternative, less rapidly cued response. We will call this the parallel competitive account, borrowing a label used by Evans (2007b) to describe Sloman’s (1996) associative and rule-based dual-process model. However, in contrast to Sloman, we do not attribute rule-based and belief-based reasoning to different systems; instead, we assume that each form of reasoning depends upon both Type 1 and Type 2 processes. For example, the conclusion to a simple argument may be generated rapidly by Type 1 processes as soon as the premise information is available, but a response must then be generated on the basis of this process that takes into account task instructions, response options, and so forth. Intentional response generation of this kind will also involve some Type 2 processing. Similarly, the evaluation of a conclusion as believable or unbelievable will involve the retrieval of relevant knowledge, which may be a Type 1 process, but then this knowledge must be converted into a similar logical form as the conclusion and these two representations will need to be compared in order for a response to be generated. This latter process must also involve some domain general Type 2 processing.

The extent to which conclusion believability interferes with logical reasoning or logical validity interferes with judgments of conclusion believability will depend upon the time it takes for each of these processes to complete. Under this account, the findings can be explained by the assumption that the arguments we employed cue a logical response before a competing response can be generated on the basis of the believability of the conclusion. Under belief instructions, the logical response must be inhibited, while the competing belief-based response is generated. This explains why logic judgments are immune to conclusion believability, but the converse does not hold when a believability judgment is required. Of course, the degree of interference observed under
belief or logic instructions will depend upon the complexity of the processes required. It is possible, for example, that the complex syllogistic arguments that are typically used in many studies of belief bias require more complex Type 2 processing than the simpler problems used here. In this case, it is quite possible that conflict will have more impact on logical than belief-based reasoning because a belief-based response may be available before processing of logical structure is completed. Although the full details of an account of this kind need to be developed and the predictions that arise need to be tested, at the very least the present findings suggest that it is overly simplistic to associate beliefs with default responses and judgments of logical validity with effortful cognition. The reality is much more complex and suggests a major re-evaluation of dual-process theories of reasoning and judgment is required.

References


Call for Papers: Special Section on Theory and Data in Categorization: Integrating Computational, Behavioral, and Cognitive Neuroscience Approaches

The Journal of Experimental Psychology: Learning, Memory, and Cognition (JEP:LMC) invites manuscripts for a special section on approaches to categorization, to be compiled by guest editors Stephan Lewandowsky and Thomas Palmeri working together with journal Associate Editor Michael Waldmann.

The goal of the special section is to showcase high-quality research that brings together behavioral, computational, mathematical, neuropsychological, and neuroimaging approaches to understanding the processes underlying category learning. There has been some divergence between approaches recently, with computational-mathematical models emphasizing the unity of category-learning processes while neuropsychological models emphasize the distinction between multiple underlying memory systems. We are seeking articles that integrate cognitive neuroscience findings in designing models or interpreting results, and behavioral studies and modeling results that constrain neuroscientific theories of categorization. In addition to empirical papers, focused review articles that highlight the significance of cognitive neuroscience approaches to cognitive theory—and/or the importance of behavioral data and computational models on constraining neuroscience approaches—are also appropriate.

The submission deadline is **June 1st, 2011**. The main text of each manuscript, exclusive of figures, tables, references, or appendixes, should not exceed 35 double-spaced pages (approximately 7,500 words). Initial inquiries regarding the special section may be sent to Stephan Lewandowsky (stephan.lewandowsky@uwa.edu.au), Tom Palmeri (thomas.j.palmeri@Vanderbilt.Edu), or Michael Waldmann (michael.waldmann@bio.uni-goettingen.de).

Papers should be submitted through the regular submission portal for JEP: LMC (http://www.apa.org/pubs/journals/xlm/submission.html) with a cover letter indicating that the paper is to be considered for the special section. For instructions to authors and other detailed submission information, see the journal Web site at http://www.apa.org/pubs/journals/xlm.