Cognition, 45 (1992) 257–284

The source of belief bias effects in syllogistic reasoning

Stephen E. Newstead
Department of Psychology, University of Plymouth, Plymouth PL4 8AA, UK

Paul Pollard
Department of Psychology, University of Central Lancashire, Preston PR1 2TQ, UK

Jonathan St.B.T. Evans
Department of Psychology, University of Plymouth, Plymouth PL4 8AA, UK

Julie L. Allen
Department of Psychology, Trinity and All Saints College, Leeds LS18 5HD, UK

Received July 26, 1991, final revision accepted May 25, 1992

Abstract

Newstead, S.E., Pollard, P., Evans, J. St.B.T. and Allen, J.L., 1992. The source of belief bias effects in syllogistic reasoning. Cognition, 45: 257–284.

In studies of the belief bias effect in syllogistic reasoning, an interaction between logical validity and the believability of the conclusion has been found; in essence, logic has a larger effect on unbelievable than on believable conclusions. Two main explanations have been proposed for this finding. The selective scrutiny account claims that people focus on the conclusion and only engage in logical processing if this is found to be unbelievable; while the misinterpreted necessity account claims that subjects misunderstand what is meant by logical necessity and respond on the basis of believability when indeterminate syllogisms are presented. Experiments 1 and 2 compared the predictions of these two theories by examining whether the interaction would disappear if only determinate syllogisms were used. It did, thus providing strong support for the misinterpreted necessity explanation. However, the results are also consistent with a version of the mental models theory, and so

Correspondence to: S.E. Newstead, Department of Psychology, University of Plymouth, Plymouth, PL4 8AA, UK.

Experiment 3 was carried out to compare these two explanations. The mental models theory received strong support, as it did also in the follow-up Experiments 4 and 5. It is concluded that people try to construct a mental model of the premises but, if there is a believable conclusion consistent with the first model they produce, then they fail to construct alternative models.

Introduction

A well-established finding in the psychology of reasoning is the belief bias effect: the finding that people are more likely to accept the conclusion to a syllogism if they believe it than if they disbelieve it, irrespective of its actual logical validity. This effect has a long history (Feather, 1964; Gordon, 1953; Henle & Michael, 1956; Janis & Frick, 1943; Kaufman & Goldstein, 1967; Lefford, 1946; Morgan & Morton, 1944; Thouless, 1959), and although some of these early studies have methodological flaws, the effect remains even when these are removed (Evans, Barston, & Pollard, 1983).

Despite the fact that the effect has been known about for all these years, and despite the extensive research that has been conducted into the phenomenon, we still know little about its source. We do not know whether it arises primarily from misinterpretation of the premises, as has been claimed by Revlin (Revlin & Leirer, 1978; Revlin, Leirer, Yopp, & Yopp, 1980); whether it arises through errors in the inferential process, as has been claimed by Oakhill and Johnson-Laird (1985); or whether it is a simple response bias effect, as suggested by Evans et al. (1983). The aim of this paper is to present experimental evidence which can help to distinguish between these various possibilities.

The essence of Revlin's claim is that believability can affect our tendency to convert the premises and thus influence the conclusion that we draw. Revlin et al. (1980) give the following syllogism as an example:

All Russians are Bolsheviks
Some Bolsheviks are undemocratic people

There is in fact no valid conclusion linking undemocratic people and Russians, but if subjects convert the first premise, believing "All Bolsheviks are Russians" to be true, then the conclusion "Some undemocratic people are Russian" follows logically. The point to note, however, is that the effect stems not from the believability of the conclusion itself but from the way in which the premises are interpreted.

Revlin et al. (1980) report two experiments designed to test their explanation. The first demonstrated that subjects' ratings of the acceptability of the converse of the premises correlated with their tendency to make the errors predicted by

conversion theory. The second attempted to show that belief bias would be eliminated if conversion was ruled out as an explanation. Both experiments can be criticized. The first used a rather strange measure of conversion, assuming that subjects were converters if they said that statements were true in their converted form for at least 60% of the entities in question. Thus subjects would be classed as converters of the statement given in the previous paragraph if they agreed to the claim that at least 60% of Bolsheviks are Russians – a questionable definition of what is meant by conversion! Their second experiment has been criticized on a number of occasions, for example for failing to control for atmosphere effects, for using primarily valid syllogisms and for confounding truth and validity (see Evans et al., 1983). Perhaps the most important aspect of this experiment, however, is that it actually found a highly significant belief bias effect even when conversion was controlled for.

Thus we can be fairly sure that belief bias is not purely an interpretational effect, but this leaves open the question as to whether interpretational factors play any role at all. Revlin et al.'s second experiment is consistent with this claim but, because of its flaws, cannot be claimed to prove it. Other research has demonstrated an apparently undiminished belief bias effect even when conversion is controlled for (Evans et al., 1983; Oakhill & Johnson-Laird, 1985). It seems reasonable to conclude that conversion errors play, at best, only a minor role in the explanation of belief bias effects.

The claim that belief bias has its effects at the inferential stage of processing is usually associated with the mental models approach to the explanation of reasoning (see Johnson-Laird & Byrne, 1991, for a recent exposition of this theory). The theory claims that syllogistic reasoning, like other kinds of reasoning, involves the construction of internal representations which are mental analogues of the situation described in the premises. Having constructed such a model, reasoners then see what conclusion, if any, is consistent with the model. They then try to falsify this conclusion by devising alternative models which are consistent with the premises but for which the conclusion does not hold. It is at this stage that believability is assumed to have its effect. If the initial model leads to a conclusion which is highly believable, then subjects will be less likely to construct alternative models than they will if the initial model corresponds to an unbelievable state of affairs.

Predictions based on this theory have been put to the test on a number of occasions. Oakhill and Johnson-Laird (1985) showed that belief bias effects could be obtained on a task in which subjects produced their own conclusions from two given premises as well as in the traditional task in which subjects evaluate given conclusions. Such a finding is difficult to explain using other theories; conversion was controlled for in this study, and response bias theories have problems since there is no conclusion present to bias the processing. The Oakhill and Johnson-Laird finding has been replicated by Markovitz and Nantel (1989).

A more direct test of the theory was provided by Oakhill, Johnson-Laird, and Garnham (1989). They investigated whether the believability of "suggested" conclusions (i.e., conclusions that were consistent with at least one of the alternative models) determined the extent of the belief bias effect. They used syllogisms such as the following:

Some of the houseowners are married
None of the houseowners is a bachelor

The valid conclusion that can be drawn is: "Some of the married people are not bachelors", but there are other believable conclusions which are compatible with at least one of the mental models which can be constructed for this syllogism, for instance: "None of the bachelors is married". Hence this is a conclusion that subjects might accept, erroneously, and fail to search for other models. On the other hand, when the alternative models lead to suggested conclusions which correspond to unbelievable states of affairs, subjects should be less inclined to consider them. This means that subjects should be more likely to make errors on the former type of problem, where there is a believable suggested conclusion, than on the latter where there is not.

The results were somewhat ambiguous. The predicted effects were obtained, but only for invalid syllogisms. For valid syllogisms such as those just presented, there was no effect of the believability of alternative conclusions, regardless of whether a construction or an evaluation task was used. However, the predicted effects were obtained with syllogisms for which there was no valid conclusion; subjects were much more likely to make errors when there was a believable conclusion that was consistent with one of the alternative models. Contrary to the authors' predictions, significant effects of believability were obtained on syllogisms where there was only one model to be constructed. Clearly the explanation in terms of alternative models simply cannot be applied in this situation; Oakhill et al. conclude that their explanation is only part of the story, and that there must also be some "conclusion filtering" taking place. What this means is that subjects evaluate any conclusion they reach, and if this is unbelievable they might reject it in favour of a more believable one.

The idea that belief bias is primarily a response effect is the traditional explanation. Most of the early work on the phenomenon assumed, either implicitly or explicitly, that this was the source of the effect, and it is also one of the favoured explanations of Evans et al. (1983). These explanations assume, in effect, that subjects suspend any attempt at reasoning and respond purely on the basis of the believability of the conclusion. Evans et al. present evidence to support this claim based on the thinking-aloud protocols produced by their subjects. Approximately 40% of subjects indicated that they had focused exclusively on the conclusion in reaching their decision, and did not mention the

premises at all, and another 30% focused initially on the conclusion and only afterwards looked at the premises; furthermore, these classifications correlated with the extent of the belief bias effect that subjects demonstrated. These authors acknowledge that response bias is not the only factor involved since they found a sizeable group of subjects who made a genuine attempt at reasoning from the premises to the conclusion; and they also found that even subjects who focused exclusively on the conclusion still showed a small effect of logic.

As we have already seen, Oakhill et al. (1989) have criticized response bias explanations on the grounds that they cannot explain the existence of belief bias effects when there is no conclusion present, as in the conclusion production task that they used. However, it could be argued that such effects can be explained on the assumption that subjects use pragmatic associations to generate conclusions and then submit these to selective scrutiny. Despite the problems, there are grounds for believing that response bias is at least part of the explanation.

It would appear, then, that all of the proposed sources of the belief bias effect have some evidence in their favour. The claim that conversion of the premises is the major source of the effect can be dismissed, but there is insufficient evidence at the moment to say that it makes no contribution at all. The mental models approach, which claims that believability has its effect on the construction of alternative models, is able to explain why belief bias is present when subjects are asked to construct their own conclusions, but fares less well in explaining the existence of belief bias on problems where only one model has to be constructed. Supporters of this approach have had to posit the existence of a mechanism for filtering conclusions which in many respects is similar to the idea that belief bias is a response effect – the main difference being that the mental models theory assumes that the believability of the conclusion has its influence after an attempt at reasoning, while response bias theorists claim that no attempt at reasoning takes place. And while response bias theory has problems with some aspects of the data, there are others that it explains rather well.

One finding that has so far resisted explanation, and which might throw some light on this issue, is the interaction between logic and belief. This is the finding that the effects of belief are more marked on invalid syllogisms than on valid ones. This interaction seems to have been first reported by Kaufman and Goldstein (1967) and also seems to have been present in the data of Oakhill et al. (1989), who found effects of believability only on indeterminate (i.e., invalid) multiple model problems. However, the most compelling illustration of the effect is in the work of Evans et al. (1983), and it is from there that we will take our example. A summary of their findings is presented in Table 1; as can be seen in this table, the difference between acceptance of believable and unbelievable conclusions is much lower for valid conclusions (33%) than it is for invalid conclusions (61%).

The explanation of this interaction which is favoured by Evans et al. is a

Table 1. *Overall percentage of acceptance of conclusions (across the three experiments) of Evans et al. (1983). Data analysed by logical validity and believability (Compiled from data in their Tables 2, 4 and 6, weighted by the N in each experiment)*

	Believable	Unbelievable
Valid	89	56
Invalid	71	10

response bias one. According to this, subjects examine the conclusion and, if it is believable, simply accept it without further ado. On the other hand, if the conclusion is unbelievable, they will then attempt to analyse the logic of the problem to see if the conclusion is valid or not. Thus logic only comes into play when the conclusion is unbelievable, thus producing the observed interaction. This explanation has been subsequently termed the selective scrutiny model (Barston, 1986; Evans, 1989).

The only other explanation of the interaction that has appeared in the literature, again proposed by Evans et al. (1983), is the misinterpreted necessity account (Barston, 1986; Evans, 1989). This is based on the arguments of Dickstein (1980, 1981) and claims that subjects fail to understand what is meant by logical necessity. Subjects are typically asked whether conclusions *necessarily* follow from the premises, which means that they should produce a conclusion only if it *must* follow from the premises. Otherwise, they should respond that nothing follows, even if one or more of the conclusions is quite possible given the premises. In essence this means that, although there is only one type of valid argument, there are two types of invalid argument: those for which the premises show the conclusion to be definitely wrong and those for which the premises give insufficient information to judge whether the conclusion would hold or not. If both premises were true, such conclusions could still be either true or false. Dickstein has argued that subjects may often misunderstand the meaning of necessity on such arguments, since it might not be intuitively obvious to them that the correct answer is that no conclusion follows when the result of their logical analysis suggests that some conclusions *might* be true. As all the invalid arguments used by Evans et al. were of this type the Dickstein argument clearly provides a simple explanation for the interaction: in the absence of a definitive result of their logical analysis on invalid problems, subjects are more likely to base their responses on the only other cue they have available, that is, their knowledge of the world.

The selective scrutiny explanation is supported by the work of Evans and Pollard (1990), who compared the extent of belief bias observed on reasoning

problems (not categorical syllogisms) involving simple and complex logic. (Logical complexity was manipulated by varying the type of logical rule and the number of premises in the problems.) They argued that increasing complexity, which in turn increases overall error rates, should lead to more belief bias according to the misinterpreted necessity but not the selective scrutiny account. The reason for this is that the former model supposes that belief bias occurs when an attempt to derive a conclusion by reasoning fails, whereas the latter sees belief bias as reflecting a heuristic which precedes and pre-empts reasoning (see Evans, 1989, for detailed discussion). Thus the selective scrutiny account predicts that additional error due to complex arguments will be random and not belief biased. The results of Evans and Pollard's experiments were in accord with these predictions.

Neither conversion theory nor mental models theorists themselves have addressed the interaction between logic and belief. Indeed, it is difficult to see how conversion theory could even attempt to explain it. The mental models theory can provide an explanation based on the believability of the initial model, and a fuller explanation of this will be given in the discussion of Experiment 2. The present Experiments 1 and 2 were designed to compare directly the selective scrutiny and misinterpreted necessity explanations of the interaction. The misinterpreted necessity explanation assumes that, on indeterminate invalid syllogisms, subjects realize (quite correctly) that no conclusion follows necessarily from the premises. However, they fail to realize that this should, given the definition of necessity, lead them to respond that no valid conclusion follows, and instead fall back on their prior knowledge when forced to make a decision. It follows from this that if all the invalid arguments used are determinately false (i.e., lead to conclusions that are definitely wrong rather than possibly wrong and possibly right), then the interaction between logic and belief should disappear. Accordingly, Experiment 1 used only determinate syllogisms.

EXPERIMENT 1

In order to produce problems that lead to determinately true or false conclusions of the same form, it is necessary to use two universal premises (which make statements about "all" the members of a class). Consider the conclusions to the following two (symbolic) arguments:

All A are B	All A are B
<u>All B are C</u>	<u>No B are C</u>
Therefore, All A are C	Therefore, No A are C
OR No A are C	OR All A are C

In each case, the first conclusion is logically valid and the second conclusion is invalid and determinately wrong. All of the problems in the present study used one of the above forms. By varying the believability of the conclusions, the four types of problems used by Evans et al. were created (i.e., valid + believable; valid + unbelievable; invalid + believable; and invalid + unbelievable); in the present study, however, the invalid conclusions were all determinately wrong.

As content in these problems, it was decided to use types and classifications of the animal kingdom. This content yields clearly believable and unbelievable statements. Arguments leading to both affirmative and negative conclusions were used to control for any effect of negation on belief. Problems led to either believable or unbelievable affirmative conclusions ("All sparrows are birds", "All ants are fish") or to believable or unbelievable negative conclusions (e.g., "No ants are fish", "No sparrows are birds"). Using this material, believable statements are definitionally true and unbelievable ones are definitionally false.

To avoid any belief effects arising in the premises, the middle term used was always a spurious (invented) category name (a method originally employed by Revlin and Leirer, 1978). A full example of a true affirmative problem shows how these were used:

All sparrows are haemopheds
All haemopheds are birds
Therefore, All sparrows are birds

Methods

Design

Subjects received eight problems in random order. For half the problems the conclusion was valid and for the other four problems the conclusion was logically invalid (and determinately wrong). Within each set of four, two conclusions were true ("believable") and two were false ("unbelievable"), yielding four cells to the design with two problems in each cell. Within each cell, one problem yielded an affirmative, and one a negative, conclusion.

Subjects

Seventy-three first-year students at the University of Plymouth participated in partial fulfilment of course requirements. None had received any formal training in logic.

Materials and procedure

All subjects received a test booklet with one page of instructions followed by eight problems, one to a page. The booklet pages were computer generated.

As the experiment was essentially devised to test the misinterpreted necessity explanation of the interaction reported by Evans et al. (1983), the instructions were devised to be as close as possible to those used in the original paper (their Experiments 1 and 2) and read as follows:

> This is an experiment to test people's reasoning ability.
>
> You will be given eight problems. On each page, you will be shown two statements and you are asked if certain conclusions (given below the statements) may be logically deduced from them. You should answer this question on the assumption that the two statements are, in fact, true. If you judge that the conclusion necessarily follows from the statements, you should answer "yes", otherwise "no".
>
> Indicate your answer by circling the appropriate word (yes or no) given below the problem. Please take your time and be sure that you have the right answer before doing so.

Content was counterbalanced across affirmative and negative conclusions, using, for instance, "All sparrows are birds" and "No sparrows are birds". However, each subject received a different content on each of the eight problems and so this counterbalancing was achieved by using two versions of the test booklet. The eight conclusions used in the two forms of the booklet are shown in Table 2. For each booklet, the computer randomly assigned one of the two conclusions of the same type to a valid, and the other to an invalid, argument. For each test booklet generated, the computer randomly assigned the eight invented middle terms to the eight problems. These middle terms were: haemopheds, bictoids, junarics, zaphods, phylones, enculions, glissomae and cryptods.

Table 2. *The eight conclusions used in the two types of test booklet in Experiment 1*

TRUE	All sparrows are birds	All rabbits are mammals
"ALL"	All trout are fish	All beetles are insects
FALSE	No rabbits are mammals	No sparrows are birds
"NO"	No beetles are insects	No trout are fish
TRUE	No pigeons are fish	No ants are fish
"NO"	No salmon are birds	No lions are insects
FALSE	All ants are fish	All pigeons are fish
"ALL"	All lions are insects	All salmon are birds

Results

Three subjects did not complete every problem and were excluded from the analyses, leaving 70 subjects. There were no differences between performance on affirmative and negative arguments. Table 3 shows the percentage frequency of acceptance of conclusions in the four cells of the design. As is clear from the table, there was a very strong effect of logical validity. Of the subjects, 68 answered more problems logically than not (sign test, $p < .001$, one-tailed) and 48 subjects made *no* logical errors at all.

Only 22 subjects made errors on any of the problems, but of these 18 made errors in favour of belief, 3 were in the opposite direction, and there was one tie. This, although limited to a minority of the subjects, is highly significant evidence of a belief bias effect (sign test, $p < .001$, one-tailed). Seventeen subjects made only belief bias errors, but the modal number of errors was one and only one subject went with belief on all eight problems.

The main point of the analysis was to test whether subjects showed more belief bias on invalid, than on valid, problems. From the means in Table 3, it can be seen that the size of this interaction is *zero* and, on a sign test, the effect was non-significantly in the opposite direction (7 vs. 10).

Discussion

As 17 subjects were categorizable as making more belief errors on either valids or invalids, it would be unreasonable to argue that the low level of belief bias observed produced an insufficiently powerful test of the interaction. This possibility is also negated by the fact that there was not even a tendency towards an interaction in the appropriate direction. This experiment, then, provides strong support for the misinterpreted necessity approach since it correctly predicts the disappearance of the interaction when determinately invalid arguments are used.

However, there are a number of ways in which the present experiment differs from those of Evans et al. (1983) which might possibly explain the discrepancies in the results. Firstly, comparison of the data for valid problems in Tables 1 and 3 suggests that the amount of belief bias in Experiment 1 is much less than that

Table 3. *Overall percentage of acceptance of conclusions in Experiment 1 as a function of logical validity and believability*

	Believable	Unbelievable
Valid	96	87
Invalid	11	2

observed by Evans et al. The problems used are ones that are known to be relatively easy ones (see, for example, Dickstein, 1978) and it is possible that this greater simplicity led to both the reduction of the belief bias effect and the disappearance of the interaction; it should be noted, however, that other studies have shown that belief bias does *not* interact with problem complexity (Evans & Pollard, 1990). Secondly, Experiment 1 used different material, and it is possible that the change in content, rather than the change in the structure of the invalid syllogisms, led to the disappearance of the interaction. Thirdly, the conclusions used in Experiment 1 were definitionally true and false rather than empirically true and false as in the material used by Evans et al. (1983). It seems probable that the difference in believability between the true and false versions of the conclusions is much greater than in the Evans et al. study, and this may have had some effect on the results. Clearly it would be desirable to repeat our experiment using problems and material more like those used by Evans et al., and thus to overcome at least the second and third problems mentioned; this was the purpose of Experiment 2.

EXPERIMENT 2

The aim of this experiment was to determine whether the results obtained in Experiment 1 could be replicated using problems and material similar to those used by Evans et al., and also controlling for the differential believability of true and false conclusions. The believability of a number of statements derived from the materials used by Evans et al. and in the present Experiment 1 was rated by 37 subjects from the University of Plymouth. Some of the statements that were rated were the same as those used originally, and others were true or false versions derived from these by changing "some . . . not" to "all" and vice versa.

It was found that the differences in believability between the true and false (i.e., believable and unbelievable) versions of the "all" and "some . . . not" versions of the Evans et al. material were very similar. On a scale from 1 to 7, where 7 corresponds to completely believable, the believable "all" statements had a mean rating of 5.64, while the unbelievable "all" statements had a mean rating of 2.28, a difference of 3.36. The believable "some . . . not" statements received a rating of 6.57, the unbelievable ones 3.40, a difference of 3.17. Hence this was the material selected for use in the present experiment. Interestingly, the differential believability of the "all" statements used in Experiment 1 was, as suspected, very high (5.64), lending some credence to the claim that there were potentially important differences between this material and that which had been used previously.

Half the subjects in this experiment received problems identical to those given

to subjects in the original study by Evans et al., which we will call the "some . . . not" problems; the other half received syllogisms such as those used in Experiment 1 but containing the same content as used by Evans et al. (these are "all" problems). Predictions for the experiment are clear-cut. As the "some . . . not" problems constitute a replication of the Evans et al. experiments, an interaction is predicted such that there is more belief bias on invalids. If this is a function of the use of conclusions which are indeterminately invalid, then there will be no such interaction on "all" problems which use determinately invalid conclusions, replicating the findings of Experiment 1. Conversely, if the disappearance of the interaction in Experiment 1 was due to the content used, there will be a similar interaction on both types of problems.

Methods

Design

Subjects received either "some . . . not" or "all" problems; in either case, they received four problems (valid/invalid × true/false).

Subjects

Sixty-one first-year undergraduate psychology students from the University of Plymouth participated in partial fulfilment of course requirements. None had received formal training in logic.

Materials

There were four "some . . . not" syllogisms, leading to valid believable, invalid believable, valid unbelievable and invalid unbelievable conclusions. To illustrate, the two arguments leading to believable conclusions are shown below:

Valid	*Invalid*
Some rich people are hard workers	No nutritional things are inexpensive things
No hard workers are millionaires	Some inexpensive things are vitamin tablets
Therefore, Some rich people are not millionaires	Therefore, Some nutritional things are not vitamin tablets

The unbelievable versions of these used the same type of syllogism but with conclusions rated as unbelievable.

Similarly, there were four "all" syllogisms, and again the two leading to believable conclusions are presented:

Valid	*Invalid*
All cigarettes are expensive things	All police dogs are vicious dogs
All expensive things are addictive things	No vicious dogs are highly trained
Therefore, All cigarettes are addictive things	Therefore, All police dogs are highly trained

The unbelievable versions of these used the same syllogistic form but with conclusions rated as unbelievable. These same four contents were used for each structure. Computer-generated booklets were produced, each containing four problems. Allocation of the two true and the two false conclusions to valid or invalid arguments was randomized.

It should be noted that the believability of the premises was not controlled; however, since this was the same for both groups of subjects, this is unlikely to have biased the results.

Procedure

Subjects received booklets in the same way as in Experiment 1, with identical instructions, followed by four problems. Since the content was the same for "all" and "some . . . not" problems, a between-subjects design was used, with 30 subjects answering "all" problems and 31 answering "some . . . not" problems.

Results

The results for the "some . . . not" problems are shown in the first half of Table 4. More subjects made errors in the direction of belief than against (sign test, 15/4 with 12 ties, $p < .01$, one-tailed) and they showed this bias more on invalid problems (sign test, 13/4 with 14 ties, $p < .025$, one-tailed).

The results for "all" syllogisms are also shown in Table 4. As in Experiment 1, there was a high frequency of correct responding although significantly more subjects made more pro- than anti-belief errors (sign test, 7/1 with 22 ties, $p < .04$, one-tailed). Nineteen of the 30 subjects answered all four questions in accordance with logic. From Table 4 it can be seen that there was slightly more belief bias on *valid* problems although this was clearly non-significant on a sign test (5/4 in favour of valids).

Table 4. *Overall percentage of acceptance of conclusions in Experiment 2*

		Believable	Unbelievable	Mean
"Some . . . not"	Valid	71	58	64
	Invalid	68	26	47
	Mean	69	42	
"All"	Valid	90	73	81
	Invalid	20	10	15
	Mean	55	41	

Discussion

The picture that emerges from Experiments 1 and 2 is strongly supportive of the misinterpreted necessity explanation of the interaction between logic and belief, but much less encouraging for the selective scrutiny approach. Just as predicted by the former, the interaction disappears when conclusions that are necessarily wrong are used rather than conclusions that are indeterminately wrong. In sharp contrast to the results of earlier studies, there is clear evidence that the selective scrutiny model does not provide an adequate explanation of the interaction, since it assumes that the effects of belief occur prior to any logical processing.

As was mentioned in the Introduction, the only two explanations of the interaction between logic and belief that have been discussed in the literature are those that have been compared in Experiments 1 and 2, and clearly it is the misinterpreted necessity account that gains most support from the results of those experiments. However, other explanations are possible, and one which deserves special attention is that based on mental models. It is in fact relatively straightforward to derive an explanation of the interaction from the work of Oakhill et al. (1989). These authors propose that the main source of the belief bias effect comes from the difficulty of constructing alternative models for multiple-model syllogisms. (Multiple-model syllogisms are ones for which more than one conceptually distinct model can be constructed; see Johnson-Laird & Bara, 1984, for a more detailed explanation.) Specifically, people will be more likely to construct more than one model when the first model they construct leads to an unbelievable conclusion. This can explain the interaction between logic and belief, since it suggests that people proceed to construct more than one mental model only when the conclusion deriving from the first model is unbelievable; if the initial conclusion is believable, then it is accepted and no further processing occurs. What this means, of course, is that full processing of all the mental models can only occur with unbelievable conclusions. As can be seen in Table 1, the nature of

the interaction is such that logic has a larger effect on unbelievable conclusions, and hence this theory predicts the observed effects.

Clearly, this theory only predicts the interaction on multiple-model syllogisms. Although the introduction of the notion of conclusion filtering allows for the occurrence of the basic belief bias effect on single model syllogisms, it cannot explain an interaction since conclusion filtering should apply equally to valid and invalid syllogisms. The syllogisms used in Experiment 1 and the "all" syllogisms used in Experiment 2 are single-model syllogisms, while those used by Evans et al. (1983) and, by extension, the "some . . . not" syllogisms of Experiment 2, are multiple-model syllogisms. This means, then, that the mental models theory can provide a perfectly adequate explanation for the results obtained in Experiments 1 and 2. The results presented so far do not allow us to distinguish between the mental models theory and the misinterpreted necessity account.

In order to distinguish between these two approaches, it is necessary to look at performance on single model syllogisms which lead to indeterminately invalid conclusions. The misinterpreted necessity theory would clearly predict that this should lead to the return of the interaction. On the other hand, the mental models theory would predict no interaction; since there is only one mental model to be constructed, belief bias can only occur through conclusion filtering and, as we have seen, this does not predict any difference in performance based on the validity of the syllogism. Experiment 3 enables a test of these contrasting predictions to be made.

EXPERIMENT 3

Consider the following syllogism:

All fish are phylones
All phylones are trout
Therefore, All fish are trout

This is a syllogism for which only one mental model can be constructed (Johnson-Laird & Bara, 1984); in the most recent notation of Johnson-Laird & Byrne (1991) this would be as follows:

[[f] = p] = t
[[f] = p] = t
. . .

In this model, f = fish, p = phylones and t = trout, the square brackets indicate

that the entity is exhaustively represented and the dots indicate that entities not mentioned in the model may exist.

This syllogism produces a valid conclusion. If the conclusion is reversed, to "All trout are fish", the syllogism is obviously still a single model one, but in this case the conclusion given is not one that validly follows. The conclusion is, however, indeterminately rather than determinately invalid, since it is possible to construct situations of this type where the conclusion could be true (for example if the t's in the above model were the only t's that existed).

Experiment 3 used syllogisms like the one given above. Half of them used forward conclusions which are valid ("All fish are trout") and half used backward, invalid conclusions ("All trout are fish"). Validity is inevitably confounded with the direction of the conclusion using this design, but this is a small price to pay for having complete control over the premises. In any case, we are not primarily interested in the effects of validity but in the interaction of this effect with belief, and any such interaction effects cannot be readily explained in terms of directional or figural effects. Believability was also manipulated so that both valid and invalid conclusions could be either believable or non-believable.

Methods

Subjects

In total, 86 students were used, 44 of whom were undergraduate psychology students, 17 of whom were chiropody students and 25 of whom were social work students, all at the University of Plymouth. None had received any formal training in logic. They were run in three separate large groups.

Materials

Booklets were prepared containing eight syllogisms in random order. In each booklet there were two problems of each type: valid believable, valid unbelievable, invalid believable and invalid unbelievable. All of the conclusions involved class inclusions, which were either believable or unbelievable. The middle terms were all meaningless, being either nonsense words or letters of the alphabet.

Procedure

Subjects were run in three groups corresponding to the courses they were taking. The instructions were the same as those used in Experiments 1 and 2.

Results and discussion

The results are presented in Table 5. As can be seen, there are some differences in the results of the three subject groups but these were not significant and hence for all analyses the data from all the groups are combined.

Overall, there was a main effect of validity, with subjects accepting significantly more valid than invalid conclusions (sign test 45/8, $p < .01$). There was also a significant effect of belief, with subjects accepting more believable than unbelievable conclusions (sign test 37/7, $p < .01$). There was, however, no interaction between logic and belief. Indeed, inspection of the means reveals that the difference between valid and invalid conclusions is identical (30%) for both believable and unbelievable items.

The results of this experiment are remarkably clear-cut. There are strong effects of belief and of logic, just as there were in Experiments 1 and 2, but there is no evidence at all of an interaction between these variables. This is precisely what would have been predicted by the mental models theory, and hence this theory is clearly supported. However, the results do not conform to the predictions of the misinterpreted necessity theory; since the invalid conclusions were indeterminate, this theory would lead one to expect a return of the interaction between logic and belief. Not only was this interaction not obtained, there was not even a hint of a trend in the expected direction.

One problem with the evidence that we have reported so far in support of the mental models approach is that it is based largely on negative predictions. In each of the experiments reported so far, the main predictions of the theory have been of an absence of the interaction between logic and belief with certain kinds of syllogisms. In Experiment 4 we tested a positive prediction – that the interaction

Table 5. *Percentage of conclusions accepted as a function of logic, belief and subject type in Experiment 3*

	Valid		Invalid	
	Believable	Unbelievable	Believable	Unbelievable
Psychologists ($n = 44$)	95	75	61	41
Chiropodists ($n = 17$)	88	82	52	29
Social workers ($n = 25$)	90	46	70	40
Overall ($n = 86$)	92	68	62	38

would return if we used multiple-model syllogisms. Hence the experiment used multiple-model syllogisms which could be either valid or invalid and lead to either believable or unbelievable conclusions.

EXPERIMENT 4

In order to create valid multiple-model syllogisms we changed the materials used in Experiment 1 such that particular ("some . . . not") conclusions were produced rather than the universal conclusions used in Experiment 1 (see Table 2). Thus conclusions such as "Some beetles are not insects" were produced in the present experiment where the original had been "All beetles are insects". The valid syllogisms were all of the kind: "Some A are B; No B are C; Therefore, Some A are not C". The invalid syllogisms were all of the kind: "No A are B; Some B are C; Therefore, Some A are not C". The believability of these statements had been assessed in the preliminary study to Experiment 2, and we know that there is a large difference between them. We also know that these differences are of the same order as those used by Evans et al. (1983), and so similar effects should be obtainable.

Methods

Subjects

The subjects used in this study were 47 undergraduate psychologists from the University of Plymouth who participated as a course requirement. None had received formal training in logic.

Materials

Two types of booklet were prepared. These each contained four syllogisms, one of each kind (i.e., valid believable, valid unbelievable, invalid believable and invalid unbelievable). The conclusions to the syllogisms were the same in each booklet: "Some beetles are not insects"; "Some pigeons are not fish"; "Some rabbits are not mammals"; Some salmon are not birds". The difference between the booklets was that valid conclusions in the first version became invalid in the second and vice versa.

Procedure

The subjects were run in groups of varying sizes. Twenty-four of the subjects received the first version of the booklet, 23 the second version. The instructions were the same as those used in the previous experiments.

Results and discussion

As can be seen from Table 6, the results are a little unusual compared with those obtained in the first three experiments. There is a significant effect of logic (sign test 31/5, $p < .001$), with valid arguments producing more acceptances than invalid ones. However, for the first time in this series of experiments there was no effect of belief (sign test 14/10, $p > .05$). The interaction between logic and belief was, however, significant (sign test 17/7, $p < .05$). This latter finding results from the fact that there was a significant effect of believability for invalid arguments but no effect at all for valid arguments (in fact, the effect actually went, non-significantly, in the opposite direction).

From the point of view of the main prediction the results are clear: the interaction between logic and belief returns when multiple-model rather than single-model syllogisms are used. This is in many ways a counterintuitive prediction, and one that is not made by any of the other theories that have been proposed to explain the interaction between logic and belief. Hence it seems reasonable to conclude that the mental models theory provides the best available explanation of the interaction.

The initially puzzling finding of an absence of an overall effect of belief in this experiment may be less surprising than it seems. Oakhill et al. (1989) failed to find an effect of belief on what they termed determinate multiple-model syllogisms. In our terminology these are multiple-model valid syllogisms – and this is precisely where we too failed to obtain an effect. Oakhill et al. were unable to suggest a plausible explanation for this finding. However, it is worth noting that these multiple-model valid syllogisms share a number of interesting features in common. They are all syllogisms for which the correct answer is "some . . . not",

Table 6. *Response frequencies (in percentages) for Experiment 4*

	Believable	Unbelievable	Mean
Valid	64	70	67
Invalid	36	17	27
Mean	50	44	

and it does seem from an examination of the literature that these are inordinately difficult. We do not know why this excessive difficulty should lead to less belief bias; indeed intuitively one might have expected exactly the opposite. Nevertheless there is clearly something different about these syllogisms that leads subjects to respond to them differently than to other syllogisms. Like Oakhill et al., however, we are unable to say what this is.

The version of the mental models theory for the explanation of belief bias effects that is being advocated here leads to other testable predictions. It is claimed that the interaction occurs because subjects fail to search for alternative, falsifying models in multiple-model syllogisms when the first model they construct leads to a believable conclusion. Hence it would be expected that the interaction would be less likely to occur if subjects can be persuaded to search for alternative models. One possible way of achieving this is through the use of instructions which stress the importance of logical necessity, pointing out that conclusions should be accepted only if they definitely follow from the premises rather than simply being possible conclusions in the light of the premises. Such instructions should make it less likely that subjects will terminate processing when they have found a believable but possible rather than necessary conclusion. This manipulation was used in Experiment 5.

This experiment also investigated the generalizability of the findings obtained in the previous experiments. Experiments 1–4 all used similar types of syllogisms: in all cases the premises were in AB–BC form, where B refers to the middle term. Experiment 5 tested for the existence of the interaction in multiple-model syllogisms of the form AB–CB. (This is the form used by Evans et al., 1983.) In addition, Experiments 1–4 used a limited range of experimental materials: all used either the material used by Evans et al. (1983) or material containing definitionally true statements linked by nonsense terms. Experiment 5 tested for the interaction in material that has not been used before and in which the middle term was not a nonsense expression.

EXPERIMENT 5

The principal aim of Experiment 5 was to compare the magnitude of the interaction between logic and belief under standard instructions (similar to those used in earlier experiments) and augmented instructions, which stressed that subjects should accept conclusions only if they necessarily followed from the premises, not if they were merely possible conclusions. This study also used neutral conclusions as well as those which are either believable or unbelievable. Previous research has produced inconclusive results using such material. Evans and Pollard (1990) found that only unbelievable conclusions differed significantly from neutral conclusions in their acceptability, suggesting that the main source of

belief bias effects may be in the rejection of unbelievable conclusions. However, Evans and Perry (1990), in a study of children's reasoning, found highly significant evidence of both positive and negative belief biases. The present study should help throw some light on this question.

Methods

Materials

The logical structure of the syllogisms used is shown in Table 7. These are all multiple-model syllogisms with premises of the form AB–CB. Half of the subjects received syllogisms in forms which involved C–A conclusions and half received forms leading to A–C conclusions. A group of subjects drawn from the same population as the experiment proper, but not participants in it, rated a set of potential conclusions chosen to provide highly believable, highly unbelievable and neutral conclusions. The ratings were made on a scale from −3 (completely believable) to +3 (completely unbelievable). The mean rating of the believable conclusions was +2.52, of the unbelievable conclusions −2.42 and of the neutral conclusions +0.48.

An example of a syllogism in the C–A direction which was valid and unbelievable is the following:

No animals are inhabitants of the island
Some tigers are inhabitants of the island
Therefore, Some tigers are not animals

Design

Subjects were divided into two groups, one of which received standard instructions whilst the other received augmented instructions. Each subject evaluated six syllogisms, three of which had a valid conclusion and three an invalid conclusion. For both valid and invalid conclusions, one was believable, one was unbelievable and one neutral and none were from the same content pair. Combination of problem type and problem content was counterbalanced across subjects so that each content occurred an equal number of times with each problem type. Problems were presented in booklet form with presentation order randomized for each subject. Responses were written in a space provided below each conclusion. Each problem appeared on a separate page.

Table 7. *Syllogisms used in Experiment 5*

Valid syllogisms:	
C–A	No A are B Some C are B Therefore, Some C are not A
A–C	Some A are B No C are B Therefore, Some A are not C
Invalid syllogisms:	
C–A	Some A are B No C are B Therefore, Some C are not A
A–C	No A are B Some C are B Therefore, Some A are not C

Subjects

Forty-eight undergraduate psychology students at the University of Plymouth took part in partial fulfilment of a course credit requirement. None had received formal training in logic.

Procedure

Subjects were run in groups of four. Each subject was given unlimited time to evaluate the validity of six syllogisms. Standard instructions were a slightly expanded version of those used in Experiments 1–4. Augmented instructions contained an additional passage outlining the principle of logical necessity with a short reminder at the end. Since the manipulation of instructions is so central to the present study, the instructions are reproduced in full. Sections in square brackets were given to the augmented instruction group only.

> This experiment is designed to find out how people solve logical problems. In the booklet which you have been given there are 6 logical reasoning problems. Your task is to decide whether the conclusion given below each problem follows logically from the information given in that problem. You must assume that all the information which you are given is true; this is very important. If, and only if, you judge that a given conclusion logically follows from the information given you should write "YES" in the space below the conclusion on that page. If you think that the given conclusion does not necessarily follow from the information given you should write "NO". [Please note that according to the rules of deductive reasoning, you can only endorse a conclusion if it

definitely follows from the information given. A conclusion that is merely possible, but not necessitated by the premises is not acceptable. Thus, if you judge that the information given is insufficient and you are not absolutely sure that the conclusion follows you must reject it and answer "NO".] Please take your time and be certain that you have the logically correct answer before stating it. If you have any questions, please ask them now as the experimenter cannot answer any questions once you have begun the experiment.

Please keep these instructions if front of you in case you need to refer to them later on. [REMEMBER, IF AND ONLY IF YOU JUDGE THAT A GIVEN CONCLUSION LOGICALLY FOLLOWS FROM THE INFORMATION GIVEN SHOULD YOU ANSWER "YES", OTHERWISE "NO".]

Please do not turn back and forth from one problem to another once you have started. You must not make notes or draw diagrams of any kind to help you in this task. Thank you very much for participating.

Results and discussion

Table 8 sets out the response patterns found for the standard and augmented instruction groups. Sign tests on problems with believable or unbelievable conclusions yielded the following results. For both instruction groups there was a substantial effect of logic ($p < .001$ in both cases). Significant effects of belief ($p = .011$) and of a belief by validity interaction ($p = .011$) were found for the standard instruction group only.

While the effect of belief is significant overall, it is notable that, as in Experiment 4, there is no effect of belief on valid problems. In fact, in both experiments there was, overall, more acceptance of unbelievable than believable conclusions for valid problems. We have no easy explanation for this finding, especially as it conflicts markedly with the findings of Evans et al. (1983) and, to a lesser extent, of the first three experiments in the present study.

Since there seem to be no effects of belief on valid problems in this study, our analysis of neutral material was restricted to invalid syllogisms. Under standard instructions, neutral conclusions had an acceptance rate intermediate between those of believable and unbelievable conclusions. The difference was, however,

Table 8. *Percentage frequencies of subjects accepting conclusions in Experiment 5 ($n = 24$ in each instruction group)*

		Believable	Unbelievable	Neutral
Standard	Valid	75	75	83
instructions	Invalid	50	0	29
Augmented	Valid	71	79	88
instructions	Invalid	17	4	25

significant only for unbelievable versus neutral problems ($p < .01$, sign test). With augmented instructions, acceptance rates for unbelievable problems were again significantly lower than for neutral problems ($p < .05$), and in this case the acceptance of neutrals was (non-significantly) *higher* than for believables. This provides support for the findings of Evans and Pollard (1990), and can be seen as favouring the claim that belief bias effects are primarily negative, resulting from the rejection of unbelievable conclusions. The implications of this finding will be considered shortly.

These results provide strong confirmation for the mental models explanation of the interaction between belief and logic; the interaction disappeared under instructions which stressed logical validity, precisely as predicted by this theory. Furthermore, the present experiment used completely new material and a different syllogistic form to that used in the earlier experiments in this paper, and the fact that the interaction was again obtained with standard instructions provides strong evidence for the generality of the earlier findings.

It is also informative to look at the performance of the two other main theories which have been put forward to explain the interaction – selective scrutiny and misinterpreted necessity – in explaining the results of Experiment 5. The predictions of the selective scrutiny model seem to be disconfirmed. According to this account, logic instructions should have an effect on overall performance but should not affect the interaction. Belief bias is assumed to be a heuristic which precedes any attempt at reasoning, and hence logic instructions should have no effect on the interaction. The finding that augmented instructions did affect the interaction clearly contradicts the predictions of this theory.

The misinterpreted necessity theory fares rather better. The interaction is assumed to stem from a failure to appreciate logical necessity, and the fact that the augmented instructions stress this would clearly lead to the prediction that these instructions should reduce the interaction – which, of course, they did. There is, however, one problem for this theory in the present results. This theory would surely predict that the augmented instructions should also have an effect on the neutral problems, since these instructions should lead to a reduction in the acceptance of invalid conclusions. As can be seen from Table 8, there was only a very small and non-significant reduction, from 29% to 25% acceptance.

At first sight, this finding might also seem to run counter to the predictions of the mental models account, since this latter might also be expected to predict that augmented instructions should increase the probability of subjects searching for alternative models when the conclusion is neutral. There is, however, a small modification to the account which can explain this finding and also why the effects of believability are stronger with unbelievable than with believable conclusions.

This modification involves assuming that the search for alternative models in multiple-model syllogisms is most strongly determined by the existence of an *unbelievable* conclusion. When an unbelievable conclusion is found after the

construction of the first model, subjects actively search for alternative models. If, however, the conclusion is believable or neutral, subjects frequently stop the search at that stage. Subjects are probably less inclined to search for alternative models when the conclusion is believable rather than neutral, but this effect is not a particularly strong one (and in fact failed to achieve significance in both this study and that of Evans & Pollard, 1990). This explains the findings on invalid problems under standard instructions in Table 8. The augmented instructions eliminate the small advantage that believable conclusions had over neutral ones, presumably by persuading subjects to search for alternative models for these. However, instructions alone are insufficient to completely eliminate the acceptance of believable or neutral invalid conclusions; only when these are made completely unbelievable does this seem to occur, at least in the present study.

The misinterpreted necessity model cannot be rescued by making a similar assumption. The problem for this theory is that logical processing is assumed to take place *before* effects of belief come into operation. Thus while augmented instructions can reduce the overall effects of believability on invalid problems, this effect should be the same for all types of problems. The finding that the augmented instructions affected believable but not neutral conclusions seems to run counter to the predictions of this model.

GENERAL DISCUSSION

There have been a number of twists and turns in the presentation of the five experiments in this paper. The first two experiments were designed to test between the two explanations that have been put forward previously of the interaction between logic and belief – the selective scrutiny account and the misinterpreted necessity account. The results clearly favoured the latter since determinately invalid syllogisms did not produce the interaction. However, in the discussion to Experiment 2 a new explanation, based on mental models, was proposed which could also explain the obtained results and so a further experiment was carried out to test between these. Somewhat to the surprise of the present authors, the results clearly favoured the mental models approach, and further experiments (Experiments 4 and 5) lent additional strong support to this.

A possible confounding factor – which it may not be possible to de-confound – is the quantifier used in the syllogisms. The single-model syllogisms in this series of experiments always used universal conclusions while the multiple-model syllogisms used particular conclusions. The reason why this confound may defy elimination, at least using categorical syllogisms, is that the only valid multiple-model syllogisms are ones which lead to "some . . . not" conclusions. The obvious way of testing for the confound is to test for the interaction in multiple-model syllogisms leading to universal conclusions, but since valid syllogisms of this kind

do not exist the experiment cannot be done. However, there seem to be sound reasons for favouring the mental models approach over any explanation involving the quantifiers used. For one thing, there seems to be no viable theoretical explanation as to why the interaction between logic and belief might itself interact with the quantifier used. On the other hand, the theory of mental models is one that has been widely investigated and which has much empirical evidence in its favour (Johnson-Laird & Byrne, 1991). Furthermore, the effects of instructions in Experiment 5 involve a separate prediction which is not confounded with the quantifier used.

We are now in a position to re-evaluate the three main explanations of belief bias effects which were discussed in the Introduction. Nothing in the results of the five experiments reported here suggests any reason for reconsidering the rather negative conclusion that was reached concerning the conversion approach. This theory seems unable to explain the interaction between logic and belief, and also seems to provide a rather inadequate explanation of the belief bias effect itself. Note, however, that our criticisms have been confined to conversion as an explanation of the phenomena of belief bias. Nothing that has been said here serves to disprove the conversion approach in general, and indeed there is strong evidence that this approach has some validity (e.g., Newstead, 1989, 1990).

The mental models approach receives strong support from the present results. The results of each of the five experiments are precisely as predicted by this theory and hence it would appear that the mental models approach provides the best explanation of belief bias effects. According to this theory, when people are given multiple-model syllogisms they attempt to construct a model consistent with the premises and, unless their initial model produces an unbelievable conclusion, they accept it without further processing. If, however, the initial conclusion is unbelievable then subjects may try to construct other models of the premises. The present results do, however, suggest a small modification to the mental models theory. The results of Experiment 5 indicate that the search for alternative models is triggered by an unbelievable conclusion, and that both believable and neutral conclusions will generally be accepted without further processing.

The predictions of the selective scrutiny account were not in the main supported by the present findings. According to this account one would expect the interaction to occur irrespective of whether the conclusions are determinately or indeterminately invalid, and irrespective of whether the syllogisms are single or multiple-model ones; the present results clearly indicate that the interaction is affected by such factors. It would, however, be premature to discard completely the *concept* of selective scrutiny, and indeed it is an integral part of the mental models account that the results seem to favour. The proposal by mental models theorists that subjects are less likely to seek counterexamples for putative believable conclusions is in itself a selective scrutiny argument.

However, the selective scrutiny account does differ from the mental models account in a crucial respect: it assumes that believable conclusions are accepted due to a heuristic which precedes and pre-empts an attempt at reasoning, whereas the mental models account assumes that some reasoning (i.e., model formation and derivation of at least one putative conclusion) occurs prior to selective scrutiny in the search for counterexamples. Clearly, the present results support the claim that some attempt at reasoning occurs prior to selective scrutiny coming into operation. Hence, although the concept of selective scrutiny may play a role, the selective scrutiny account, as propounded in the Introduction, appears to be disconfirmed.

One finding of Oakhill et al. (1989) that has been confirmed in the current experiments is that belief bias effects occur on single-model syllogisms. The selective scrutiny account provides a perfectly viable explanation of this phenomenon, but so too does the notion of conclusion filtering within the mental models approach. In fact, once again, the theories are not dissimilar to each other. Selective scrutiny claims that subjects filter out conclusions prior to logical processing, while the mental models theory claims that filtering occurs after some analysis; but the process itself is virtually identical and, indeed, it is possible that filtering occurs at both stages.

There is another possible explanation for belief bias effects on single-model syllogisms which also deserves consideration. When subjects have constructed their original model and found this to be unbelievable, they are assumed to search for alternative, falsifying, models. With single-model syllogisms no such models can be constructed, but this does not mean that subjects will not try to find them. It is possible that in some circumstances subjects think (incorrectly) that they have found such a model and hence reject unbelievable conclusions. This explanation fits the data and also accords well with the other assumptions of the mental models approach, but in the absence of further evidence there are no grounds for preferring it over other explanations.

It should be noted at this point that although the weight of evidence is clearly behind the mental models explanation of the interaction, there are still one or two puzzling findings that the account does not readily explain. For example, Evans et al. (1983) found in their verbal protocols that many subjects seemed to inspect only the conclusions of syllogisms, which is clearly inconsistent with the claim that they should construct a mental model from the premises. It is possible that there are individual differences in the way in which people tackle syllogisms, and that not everyone attempts to construct mental models. There is indeed a strong possibility that more than one theory may be appropriate to explain performance in syllogistic reasoning tasks. However, the results of the present experiments suggest very strongly that mental models will provide an important component of an overall theory.

References

Barston, J.L. (1986). *An investigation into belief biases in reasoning*. Unpublished Ph.D. thesis, University of Plymouth.

Dickstein, L.S. (1978). The effect of figure on syllogistic reasoning. *Memory & Cognition*, *6*, 76–83.

Dickstein, L.S. (1980). Inference errors in deductive reasoning. *Bulletin of the Psychonomic Society*, *16*, 414–416.

Dickstein, L.S. (1981). The meaning of conversion in syllogistic reasoning. *Bulletin of the Psychonomic Society*, *18*, 135–138.

Evans, J.St.B.T. (1989). *Bias in human reasoning*. Hove, UK: Lawrence Erlbaum.

Evans, J.St.B.T., Barston, J.L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*, 295–306.

Evans, J.St.B.T., & Perry, T. (1990). *Belief bias in children's reasoning*. Unpublished manuscript, Department of Psychology, University of Plymouth.

Evans, J.St.B.T., & Pollard, P. (1990). Belief bias and problem complexity in deductive reasoning. In J.P. Caverni, J.M. Fabre, & M. Gonzales (Eds.), *Cognitive biases* (pp. 131–154). Amsterdam: North-Holland.

Feather, N.T. (1964). Acceptance and rejection of arguments in relation to attitude strength, critical ability and intolerance of inconsistency. *Journal of Abnormal and Social Psychology*, *69*, 127–136.

Gordon, R. (1953). Attitudes toward Russia on logical reasoning. *Journal of Social Psychology*, *37*, 103–111.

Henle, M., & Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *Journal of Social Psychology*, *44*, 115–127.

Janis, I., & Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. *Journal of Experimental Psychology*, *33*, 73–77.

Johnson-Laird, P.N., & Bara, B.G. (1984). Syllogistic inference. *Cognition*, *16*, 1–61.

Johnson-Laird, P.N., & Byrne, R. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum.

Kaufman, H., & Goldstein, S. (1967). The effects of emotional value of conclusions upon distortions in syllogistic reasoning. *Psychonomic Science*, *7*, 367–368.

Lefford, A. (1946). The influence of emotional subject matter on logical reasoning. *Journal of General Psychology*, *34*, 127–151.

Markovitz, H., & Nantel, G. (1989). The belief bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*, 11–17.

Morgan, J.J.B., & Morton, J.T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, *20*, 39–59.

Newstead, S.E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, *28*, 78–91.

Newstead, S.E. (1990). Conversion in syllogistic reasoning. In K. Gilhooly, M.T.G. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking* (Vol. 1, pp. 73–84). Chichester: Wiley.

Oakhill, J.V., & Johnson-Laird, P.N. (1985). The effects of belief on the production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology*, *37A*, 553–569.

Oakhill, J.V., Johnson-Laird, P.N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, *31*, 117–140.

Revlin, R., & Leirer, V.O. (1978). The effects of personal biases on syllogistic reasoning: Rational decisions from personalised representations. In R. Revlin & R.E. Mayer (Eds.), *Human Reasoning*. Washington: Winston/Wiley.

Revlin, R., Leirer, V.O., Yopp, H., & Yopp, R. (1980). The belief bias effect in formal reasoning: The influence of knowledge on logic. *Memory & Cognition*, *8*, 584–592.

Thouless, R. (1959). The effect of prejudice on reasoning. *British Journal of Psychology*, *50*, 289–293.