

Believability and syllogistic reasoning*

JANE OAKHILL

University of Sussex

P.N. JOHNSON-LAIRD

*MRC Applied Psychology Unit,
Cambridge*

ALAN GARNHAM

University of Sussex

Received September 1987, final revision accepted November 1988

Abstract

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

In this paper we investigate the locus of believability effects in syllogistic reasoning. We identify three points in the reasoning process at which such effects could occur: the initial interpretation of premises, the examination of alternative representations of them (in all of which any valid conclusion must be true), and the “filtering” of putative conclusions. The effect of beliefs at the first of these loci is well established. In this paper we report three experiments that examine whether beliefs have an effect at the other two loci. In experiments 1 and 2 subjects drew their own conclusions from syllogisms that suggested believable or unbelievable ones. In the third experiment they evaluated conclusions that were presented to them. The data show that beliefs both affect the examination of alternative models and act as a filter on putative conclusions. We conclude by showing how some types of problem and some problem contents make the existence of alternative models more obvious than others.

Introduction

Logicians draw a distinction between the validity of an inference and the truth of its premises and conclusion. Although a false conclusion can never

*Requests for reprints should be addressed to: Jane Oakhill, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, U.K.

follow validly from premises that are all true, there are no other constraints on combinations of true and false premises and valid and invalid arguments. Furthermore, in deriving or assessing a conclusion from a set of premises, believability should not be taken into account. Nevertheless, people's beliefs influence the conclusions they draw and the way they evaluate arguments. They tend to draw or accept believable conclusions too readily and to be wary of drawing or accepting unbelievable ones. For example, beliefs can bias reasoning from syllogistic premises, both when people are asked to evaluate given conclusions (Evans, Barston, & Pollard, 1983) and when they draw conclusions for themselves (Oakhill & Johnson-Laird, 1985). The latter study showed in particular that, if the premises lead validly to a conclusion, then people are more likely to produce that conclusion if it is believable than if it is unbelievable. However, this result held only when a conclusion was either true or false by definition rather than as a matter of fact. So, unbelievable conclusions such as "Some of the actresses are not women" were produced far less frequently than believable ones such as "Some of the athletes are not healthy". When the premises did not support a valid conclusion, there was a corresponding effect: subjects were more likely to respond "no valid conclusion" if the premises suggested a *definitionally* false conclusion.

The question we wish to address in the present paper is *how* beliefs affect reasoning. One old idea is that beliefs affect how likely a premise is to be converted. For example, "All spaniels are dogs" is unlikely to be converted to "All dogs are spaniels" because almost everyone is familiar with other breeds of dogs. However, an abstract premise, such as "All A are B", or one that describes a relation between unfamiliar objects, such as "all foraminifera are rhyzopoda", may well be mistaken for the assertion that "All B are A" or "all rhyzopoda are foraminifera" (see Wilkins, 1928). However, our previous results (Oakhill & Johnson-Laird, 1985, experiment 2) show that conversion cannot explain all the effects of belief on syllogistic reasoning. This experiment showed that bias effects arise even for problems in moods in which conversion of the premises does not alter the valid conclusion.

These results suggest that beliefs can affect the process of making an inference and evaluating a putative conclusion. This idea poses a major problem in accounting for the effects of belief on reasoning. In the past, it has typically been assumed that people reason using formal rules of inference (e.g., Braine, 1978; Rips, 1983). However, it is difficult to see how beliefs could interact with such rules which, by definition, make no reference to any specific content.

We will not attempt to modify a theory of this kind, but rather we will show how a different kind of account of syllogistic reasoning can more naturally accommodate the effects of belief on reasoning. We believe that deductive reasoning depends, not on rules of logic, but on the construction and

manipulation of mental representations, or *models*, of the states of affairs described in the premises (Erickson, 1974; Guyote & Sternberg, 1981; Johnson-Laird, 1983). We will not recapitulate the theory in detail here since it has been described elsewhere (Johnson-Laird, 1983, chap. 5; Johnson-Laird & Bara, 1984a). For the present paper, the principal issue is whether a problem requires one model or more than one model. Multiple models are needed whenever there is more than one way of adding the information from one premise to a model of the other premise.

Consider the premise "All of the artists are beekeepers", which supports a model of the form:

artist = beekeeper
 artist = beekeeper
 O beekeeper
 O beekeeper

where each line represents a separate individual and the O (for optional) indicates that beekeepers who are not artists may, or may not, exist in the domain of discourse. Now, suppose that there is a second premise "All of the beekeepers are chemists". In adding this information to the model there is no choice about what to do: wherever there is a beekeeper it must be tagged as a chemist. The resulting model supports the conclusion "All of the artists are chemists". But, suppose instead that the second premise is "Some of the beekeepers are chemists". Now there is a choice about which beekeepers to tag as chemists: those that are artists, or those that are not, or some mixture of the two. Whenever there is such a choice, more than one model of the premises is possible. In this case, the different models do not support any single conclusion interrelating artists and chemists. Therefore, no conclusion is valid. In other multiple-model problems there is a valid conclusion interrelating the two end terms. The precise number of alternative models depends on the particular procedures that are assumed to be used in constructing models: Johnson-Laird and Bara (1984a) described two different sets of procedures that produce differing numbers of alternative models. What is common to both sets of procedures, however, is the set of problems that they classify as having one model—those for which there is no choice about how to add the information from the second premise. In the present study we have therefore distinguished three sorts of problem: one-model problems (which all have valid conclusions), multiple-model problems with valid conclusions (henceforth *determinate multiple-model problems*) and multiple-model problems with no valid conclusion (henceforth *indeterminate problems*).

Beliefs could have an effect at any stage of the reasoning process. Most straightforwardly, they could directly affect the initial interpretation of indi-

vidual premises. For example, knowledge of dogs that are not spaniels might affect the status of the optional elements in the representation of "All spaniels are dogs". The existence of such effects is hardly controversial (e.g., Wilkins, 1928), and we do not propose to investigate them any further. In this paper, we will be concerned with two other ways in which beliefs could affect reasoning. First, whether reasoners consider all the alternative models in multiple-model problems could be influenced by their prior beliefs. If they find a believable conclusion that is true in the model(s) they have so far constructed, they may accept that conclusion, rather than try to find another model in which it is false. Conversely, there might be a bias against accepting unbelievable conclusions without a thorough search for alternative models. Second, beliefs could act as a "filter" on putative conclusions before they are finally accepted. As a result of such filtering, reasoners might change a conclusion into one that is more believable, without further reference to models of the premises, or else they might claim that there is no valid conclusion.

The data from our previous study (Oakhill & Johnson-Laird, 1985) do not enable us to distinguish between these possible locations of the effect of beliefs on reasoning. However, the two ideas give rise to different predictions. If beliefs exert their effect on the *process* of model building (rather than as a filter on putative conclusions), they should have different effects on different types of syllogism. For one-model syllogisms they can have no effect because there are no alternative models. For multiple-model syllogisms, on the other hand, they can have an effect. If, say, the first model constructed for a multiple-model problem suggests a conclusion that is highly believable, the search for alternative models may be curtailed, and the believable, but possibly invalid conclusion accepted. But if the initial conclusion is unbelievable, the search for alternative models should continue. In principle, it should eventually lead to the correct conclusion, but a wealth of previous data shows that people find determinate multiple-model syllogisms very difficult and only rarely deduce correct conclusions from them.

If the alternative hypothesis—that believability acts as a final filter on all conclusions—is correct, belief should affect both one- and determinate multiple-model problems, since both can have believable or unbelievable conclusions.

The materials in these experiments differed from those in our previous experiments (Oakhill & Johnson-Laird, 1985) in that we manipulated the believability of the *suggested* conclusions. By "suggested", we mean conclusions that are compatible with at least one model of the premises, but not with all of them in cases where there is more than one model. For the one-model problems, there is only a single correct representation of the premises, so there are no alternative models to consider when a conclusion is unbeliev-

able. For someone reasoning logically, there is no option but to accept that conclusion. To put this point another way, for one-model problems, the suggested conclusions were the correct ones. For the determinate multiple-model problems, the correct conclusions were always believable, for example:

Some of the proud people are not humble
Some of the dishonest people are not liars

However, the suggested conclusions could be either believable or unbelievable. For example, the syllogisms that led to the believable conclusions above had, respectively, the following suggested conclusions:

None of the humble people is proud
None of the liars is dishonest

If believability affects the model construction process, subjects should tend to accept an initial (suggested) conclusion to such a problem if it is believable, and should stop the deductive process at that point. Otherwise they should go on to try to find an alternative model.

Our materials also included indeterminate syllogisms, all of which are multiple-model. These problems cannot be used to address the filtering hypothesis, since the fact that there is no valid conclusion cannot be filtered in the way that a believable or unbelievable conclusion can. Nevertheless, they can provide evidence relevant to the hypothesis that beliefs affect the process of model building and testing. For these syllogisms we again manipulated whether the set of suggested conclusions was believable or not. If believability affects the deductive process, subjects should be more likely to produce the correct response to these problems (NVC = "no valid conclusion") when the suggested conclusions are unbelievable, and more likely to produce suggested conclusions in error when they are believable.

Before describing the experiments, we will outline the construction of the materials, most of which were common to all three experiments.

Materials

Before we selected contents for the syllogisms we selected six syllogistic forms for the experiments: two one-model forms, two indeterminate forms and two determinate multiple-model forms. Across the different forms, the conclusions, either valid or suggested, were in different moods. The one-model syllogisms had valid conclusions in moods E (Figure 1, AE) and I (Figure 1, IA)—here, and elsewhere, the figures we refer to are those of Johnson-Laird and Bara (1984a) rather than the traditional ones. The indeterminate syl-

logisms had suggested conclusions in moods A (Figure 3, AA) and I (Figure 1, AI). All determinate multiple-model syllogisms, including those chosen for the experiment (Figure 4, EA and Figure 4, IE), have a valid conclusion in mood O, and suggested conclusions in moods E and O. For each of six chosen forms we used two unrelated contents.

Once the forms of the syllogisms were selected, suitable believable and unbelievable valid or suggested conclusions were chosen in a rating study. The materials for rating were complex, and will be described in detail. Our aim was to find matched sets of believable and unbelievable conclusions suggested by initial models. For example, for a one-model syllogism, matched believable and unbelievable conclusions might be:

Some of the clever people are geniuses
Some of the clever people are stupid

Given a conclusion and a syllogistic form, appropriate premises can be constructed by selecting a neutral middle term.

In all the experiments we used statements in the *of the* form (e.g., *all of the A are B* rather than *all A are B*). The use of the definite article suggests that a particular set of As are under discussion. It also indicates a presupposition that those As exist (see Johnson-Laird & Bara, 1984b, for further discussion).

We constructed 22 sets of conclusions (suggested and valid). A large proportion included both O and E conclusions and were intended for determinate multiple-model problems, since it is more difficult to obtain satisfactory ratings for sets of four conclusions than for sets of two. For these syllogisms, each of the four conclusions was compatible with at least one model, but only one was valid (i.e., compatible with all models). For the one-model syllogisms there were two conclusions in the set. Both were valid, since they were in moods E and I and, hence, convertible. If *some of the A are C* is true then so is *some of the C are A*, and similarly with the universal negative (E) conclusions. For the indeterminate syllogisms, each set also had two conclusions. These conclusions were the two possible A conclusions or the two possible I conclusions. Although other conclusions were compatible with one or other model in the case of one indeterminate syllogism (if an A conclusion holds then so does the corresponding I) these conclusions were never produced by subjects asked to generate conclusions from syllogistic premises (Johnson-Laird & Bara, 1984a).

Each set of conclusions had a believable form, and an unbelievable form. For the one-model and indeterminate syllogisms these forms were quite straightforwardly derived. The pairs of conclusions were either both believable or both unbelievable. For the determinate multiple-model syllogisms, the

valid conclusion was always intended to be believable. The suggested conclusions were intended to be either all believable or all unbelievable.

From these sets of conclusions, two lists were produced, and the believable and unbelievable versions of each set were assigned to separate lists. Thus, one list contained the following pair of conclusions:

Some of the happy people are depressed
Some of the depressed people are happy

in which both statements are unbelievable, and the other contained the corresponding believable pair:

Some of the sad people are depressed
Some of the depressed people are sad

Each list contained equal numbers of sets intended to be either believable or unbelievable.

The 16 raters were presented with *complete* sets of conclusions, as we felt that comparative assessment of related conclusions would more closely approximate what happens when subjects are solving a syllogism and evaluating putative conclusions. However, the raters were *not* given any premises. They were simply asked to rate the plausibility of the conclusions as statements, *not their plausibility as conclusions from given premises*. Eight raters received each list. Their task was to rate each statement on a scale of 1 (completely unbelievable) to 7 (completely believable). The raters were told that the statements were in sets of 2 or 4 and that, in some cases, all of the statements in a set were very similar, so that it might be appropriate to give them all a similar believability rating. It was also emphasised that, in other cases, the items in a set might require very different ratings. Raters were also told to bear in mind the relative ratings for all the statements within one set.

We derived a mean rating for each statement, and selected the most clearly and consistently rated sets of statements. We derived problems in which those statements could act as conclusions, as follows. In the case of one-model problems, the *correct* conclusions were either believable or unbelievable. They were also convertible, so both pairs in the set had to have similarly rated conclusions (both believable or both unbelievable). An example of a one-model problem in its two forms is shown below (here, and elsewhere, conclusions that follow validly from the premises are shown in capitals).

Some of the communists are golfers
All of the golfers are Marxists

SOME OF THE COMMUNISTS ARE MARXISTS (believable, mean rating 6.6)

SOME OF THE MARXISTS ARE COMMUNISTS (believable, mean rating 6.1)

Some of the communists are golfers
All of the golfers are capitalists

SOME OF THE COMMUNISTS ARE CAPITALISTS (unbelievable, 2.4)
SOME OF THE CAPITALISTS ARE COMMUNISTS (unbelievable, 2.1)

For the determinate multiple-model problems, the situation was different. The valid conclusion was always believable. We therefore chose related pairs of conclusion sets such that in one set all the statements were believable and in the other only the valid conclusion was believable. An example of such a problem in its two forms is shown below.

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

None of the married people is a bachelor (believable, 4.4)
None of the bachelors is a married person (believable, 4.3)
SOME OF THE MARRIED PEOPLE ARE NOT BACHELORS (believable, 7.0)
Some of the bachelors are not married (believable, 7.0)

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

None of the married people is a husband (unbelievable, 3.5)
None of the husbands is a married person (unbelievable, 1.0)
SOME OF THE MARRIED PEOPLE ARE NOT HUSBANDS (believable, 6.0)
Some of the husbands are not married (unbelievable, 1.0)

In the former set, all statements are reasonably believable but in the latter only one conclusion, the valid one, is.

For the indeterminate syllogisms, as with the one-model problems, we chose sets of conclusions that were either both believable or both unbelievable. The two conclusions were each compatible with one or other model of

the premises. An example of the two versions of an indeterminate problem is shown below.

All of the Frenchmen are wine drinkers
Some of the wine drinkers are gourmets

Some of the Frenchmen are gourmets (believable, 6.6)
Some of the gourmets are Frenchmen (believable, 6.4)
NO VALID CONCLUSION

All of the Frenchmen are wine drinkers
Some of the wine drinkers are Italians

Some of the Frenchmen are Italians (unbelievable, 1.6)
Some of the Italians are Frenchmen (unbelievable, 1.6)
NO VALID CONCLUSION

Experiment 1

Method

Two lists of materials were derived from 12 selected matched pairs of syllogisms. Each list contained four of each of the three types of syllogism: one-model, determinate multiple-model and indeterminate. The suggested conclusions (valid for one-model problems only) for half of the problems of each type in each list were believable, and for the other half they were unbelievable. If one list contained the believable version of a particular problem, the other list contained the unbelievable version. These lists were presented to different subjects so that each subject received each form of syllogism with two unrelated contents: one believable and one unbelievable. The syllogisms were typed, one to a page, and stapled into booklets. Each of the three types of syllogism (one-model, determinate multiple-model and indeterminate) appeared twice in each half of a booklet, once with a believable suggested conclusion, and once with an unbelievable one. Each type of syllogism occurred in two forms (e.g., one-model syllogisms appeared in Figure 1, AE and Figure 1 IA). One form was associated with a believable conclusion in the first half of the booklet, and with an unbelievable conclusion in the second half, or vice versa. The order of syllogisms within each half of the booklet was random.

As well as the 12 experimental syllogisms, each booklet also contained two filler syllogisms at the beginning, to give the subjects an easy "lead in" to the task.

The booklets were presented to 45 subjects. Of these, 12 were students at the University of Sussex, who were paid for participation and were tested individually. The other 33 subjects were students at the University of Reading, who participated in the experiment as part of a practical class, and were not paid. None of the subjects had received any training in formal logic. Roughly the same number of subjects received each of the two lists. All subjects were given both detailed written instructions and a verbal explanation of the task. They were told that they were to assume that all the statements were true and that their "conclusion should be based solely on what can be deduced with absolute certainty from the premises". In particular, they were warned against making inferences that were plausible, but not certain. Their task was to write down a single conclusion to each problem, or to state that there was "no valid conclusion". Subjects were given as long as they needed to complete the task.

Results

The data from 4 subjects (2 from Sussex and 2 from Reading) were excluded from the analysis because those subjects produced a high proportion of inadmissible responses (i.e., ones containing middle terms or in which new terms were introduced).

The results for the remaining 41 subjects are summarised in Table 1. Unacceptable responses, of which there were never more than 4/82 per condition, were excluded. Responses were unacceptable if they included the middle term instead of, or as well as, the end terms, or if the end terms were

Table 1. *Correct and suggested conclusions produced in Experiment 1 as a percentage of acceptable conclusions (see text). In every case the maximum number of acceptable conclusions was 82*

	Type of problem		
	One-model	Determinate multiple-model	Indeterminate
<i>Correct</i>			
Believable	95	14	30
Unbelievable	65	7	73
<i>Suggested</i>			
Believable	n/a (NVC = 5)	37 (NVC = 49)	67
Unbelievable	n/a (NVC = 21)	32 (NVC = 56)	13

qualified in any way. In a few cases, a problem had been missed. In an analysis of correct responses, Wilcoxon T tests showed significant effects of believability on one-model ($T = 20$, $N = 22$, $p < .01$) and indeterminate problems ($T = 20$, $N = 28$, $p < .01$). For the indeterminate problems, subjects were correct a high proportion of the time when the suggested conclusions were unbelievable, that is they only accepted these conclusions 13% of the time, but when the suggested conclusions were believable they were drawn on 67% of trials ($T = 0$, $N = 30$, $p < .01$).

For the determinate multiple-model problems there were too few cases in which the correct conclusion was produced to warrant an analysis. There was no indication that the suggested conclusions were accepted more often when they were believable ($T = 120$, $N = 23$, n.s.)—the predominant response for these problems was “no valid conclusion”. Slightly more NVC responses were produced when the suggested conclusion was unbelievable, but the difference was not significant ($T = 73.5$, $N = 20$).

Before discussing the results of this experiment in detail we will present the results of two further studies.

Experiment 2

The very small number of correct responses to the determinate multiple-model problems in Experiment 1 was surprising, since subjects in previous experiments performed quite well when they were given predominantly difficult (multiple-model) problems (Oakhill & Johnson-Laird, 1985). Perhaps the inclusion of “easy” problems produced a response bias towards NVC responses when a valid conclusion could not readily be derived. In the earlier experiments, with a high proportion of difficult problems, subjects probably realised that NVC was unlikely to be correct in the majority of cases, and tried harder to derive a conclusion. In Experiment 2, therefore, we gave the one-model problems and determinate multiple-model problems to different groups of subjects. Because the determinate problems were now so similar, we included more filler items, to introduce a wider variety of problems and conclusions and to distract the subjects from the pattern of the premises. Unfortunately, it is impossible to achieve a *very* high proportion of difficult problems while maintaining a variety of valid conclusions, because *all* valid conclusions to determinate multiple-model problems are in mood O (some ... are not ...).

Method

The materials were the same as those used in Experiment 1, with the exception of the additional fillers. These fillers were added so that each of the five possible conclusions (A, E, I, O, NVC) was valid for at least one syllogism per booklet. Booklets were made up as before from the two lists of materials, except that *two* booklets were derived from each list. One contained all the one-model problems, and the other the determinate multiple-model problems. The indeterminate problems from a particular list were all included in both booklets, and the four filler items were the same for all four types of booklet. Thus, each booklet contained 12 problems: 4 determinate, 4 indeterminate and 4 fillers. Within each half of the booklet, there were equal numbers of determinate and indeterminate problems—the order of problems within each half was randomised. The filler items always occurred in 1st, 4th, 8th and 11th positions within the booklet.

The subjects were 21 students from the University of Sussex, who took part in the experiment as part of the practical class, and were not paid. As far as was possible, equal numbers of subjects were assigned to each of the four booklet types. None of the subjects from Experiment 1 took part in this experiment, and none had had any training in logic.

Results

The results are shown in Table 2. Again, very few responses had to be excluded. The overall pattern was very similar to that of Experiment 1, al-

Table 2. *Correct and suggested conclusions produced in Experiment 2 as a percentage of acceptable conclusions (see text). The maximum number of acceptable conclusions was 22 for one-model problems, 20 for determinate multiple-model problems and 42 for indeterminate problems*

	Type of problem		
	One-model	Determinate multiple-model	Indeterminate
<i>Correct</i>			
Believable	86	6	21
Unbelievable	62	22	74
<i>Suggested</i>			
Believable	n/a (NVC = 5)	67 (NVC = 28)	72
Unbelievable	n/a (NVC = 38)	33 (NVC = 22)	8

though the difference between number of correct responses for the one-model believable and unbelievable problems was not significant ($T = 6$, $N = 7$), perhaps because of the small number of subjects (11) who were given problems of this kind. As before, there was no significant difference in the number of correct responses for the determinate multiple-model believable and unbelievable problems ($T = 0$, $N = 3$). For these problems, more believable than unbelievable suggested conclusions were produced, but the difference was not significant ($T = 3.5$, $N = 7$). Almost all of these conclusions, both in this experiment and in Experiment 1, were of the form "No A are C" or its converse.

As in Experiment 1, there were significantly more correct (i.e., NVC) responses to the indeterminate problems with unbelievable suggested conclusions than those with believable conclusions ($T = 4.5$, $N = 15$, $p < .005$, one-tailed), and significantly more believable than unbelievable conclusions were produced in error to these problems ($T = 0$, $N = 16$, $p < .005$, one-tailed).

Experiment 3

Increasing the proportion of determinate multiple-model problems, in Experiment 2, failed to increase the number of correct responses to them. We therefore decided to investigate how subjects performed when *evaluating*, rather than producing, believable and unbelievable conclusions. We expected that subjects would be better at evaluating given conclusions to such problems (cf. Evans et al.'s, 1983 vs. Oakhill & Johnson-Laird's, 1985, results). We also asked subjects to write down the conclusion they thought they could draw, if they rejected the presented conclusion.

Method

We used the same materials as in Experiment 1 and presented them with conclusions for evaluation, as follows.

Determinate problems

We presented the determinate problems with their believable or unbelievable "suggested" conclusions which, in the case of the one-model problems, were also the valid conclusions. For the multiple-model problems, the conclusions presented were always of the E form ("None of the ..."). The *valid* conclusions were not given for evaluation. For these problems, the valid conclusions were always believable, because we were interested in the effects

of suggested conclusions, and wished to keep the believability of correct conclusions constant.

The predictions were similar to those for Experiments 1 and 2 in which subjects had to produce their own conclusions. If believability affects the examination of alternative models, the believability of the presented conclusion for one-model problems should have no effect on its final evaluation since, even if lack of believability prompts a thorough search for an alternative model, there is no such model that satisfies the premises and refutes the given conclusion. However, if beliefs act as a filter they should produce an effect. In the case of multiple-model problems, a believable presented conclusion might curtail the examination of alternative models if it is true in the model first constructed. However, an unbelievable suggested conclusion could act as a cue that the examination of alternative models should proceed. We therefore expected the subjects to reject the unbelievable conclusions and, possibly, to produce the correct conclusion that is consistent with all the models.

Indeterminate problems

In this experiment, we included two types of indeterminate problem. First, there were those used in the experiments above, which were presented with either a believable or an unbelievable suggested conclusion. As with the determinate multiple-model problems, we predicted that, if believability affects the examination of alternative models, subjects should accept the believable suggested conclusions, since they are consistent with one model of the premises, but would be prompted by unbelievable conclusions to search for an alternative model, and should be more likely to reach the correct conclusion (that nothing follows). Second, in order to include some indeterminate problems for which the given conclusion ("no valid conclusion") was correct we produced four filler items, which were the same for all subjects. The filler problems were constructed so that the suggested conclusions—ones that occurred frequently as errors in Johnson-Laird and Bara's (1984a) experiments—were believable or unbelievable. We used two syllogistic forms for these filler items, examples of which are shown below:

All of the doctors are egotists
Some of the egotists are general practitioners

NO VALID CONCLUSION

(suggested believable conclusion: Some of the doctors are general practitioners)

Some of the actresses are not hikers
 All of the hikers are women

NO VALID CONCLUSION

(suggested unbelievable conclusion: Some of the actresses are not women)

Since these items were conceived as fillers, they were not constructed as systematically as the experimental items, although reference was made to the previously collected ratings. We did not have matched pairs of premises with believable and unbelievable suggested conclusions.

The booklets for this experiment were constructed in a similar way to those for Experiment 1. Of the four fillers, two were assigned to each half of the list (one in each form; one believable and one unbelievable).

Subjects

Thirty-two subjects were tested. Sixteen performed the experiment as part of a practical class at Sussex University and were not paid. The other 16 were staff and students at Sussex University who were tested individually and who were paid. The data from two subjects in the practical class were excluded, because they discussed the problems with one another. None of the subjects from Experiments 1 or 2 participated in this experiment, and none had had any training in logic.

Results

The data from the remaining 30 subjects are summarised in Table 3. Performance on the one-model problems was at ceiling level (almost all subjects got them all right, regardless of believability), and no statistical analysis was carried out. Unexpectedly, more conclusions to determinate multiple-model problems were correctly rejected when they were believable than when they were unbelievable, but the difference was not significant ($T = 36$, $N = 17$). Neither was there a significant difference in the number of NVCs produced in error to these problems ($T = 72.5$, $N = 19$). There were too few correct and suggested conclusions produced for an analysis to be feasible.

In the case of indeterminate problems, significantly more suggested conclusions were correctly rejected when they were unbelievable ($T = 17$, $N = 18$, $p < .005$), and more correct NVC conclusions were produced in this condition, although the difference was not significant ($T = 37.5$, $N = 16$).

The data from the filler items are also of some interest and are shown in Table 4.

The indeterminate filler items (presented with their correct conclusion:

Table 3. *Percentages of conclusions accepted, rejected and produced in Experiment 3*

	Type of problem		
	One-model	Determinate multiple-model	Indeterminate
<i>Suggested conclusion believable</i>			
Suggested accepted	95 (correct)	30 (incorrect)	48 (incorrect)
Correct produced	n/a	5	45
NVC wrongly produced	2	42	n/a
Other suggested produced	2	13	3
<i>Suggested conclusion unbelievable</i>			
Suggested accepted	90 (correct)	45 (incorrect)	22 (incorrect)
Correct produced	n/a	8	58
NVC wrongly produced	7	33	n/a
Other suggested produced	2	10	2

Table 4. *Percentages of conclusions accepted, rejected and produced for the (indeterminate) filler items in Experiment 3*

Suggested conclusion	Correct (NVC) accepted	Suggested produced
Believable	45	57
Unbelievable	62	13

NVC) showed a complementary pattern of results to the other indeterminate problems. The correct conclusion was more often accepted when the suggested conclusion was unbelievable, though the difference was not significant ($T = 57$, $N = 20$). More suggested conclusions were produced in error when they were believable ($T = 22.5$, $N = 17$, $p < .005$).

General discussion

One-model problems

In all three experiments there was a tendency for performance to be better on one-model problems when their conclusions were believable than when they were unbelievable. This effect was significant in Experiment 1, marginally significant in Experiment 2, where the number of subjects was less, and

was masked by a ceiling effect in Experiment 3. This result cannot be explained on the assumption that believability has its effect *solely* on the process of constructing alternative models of the premises, because these problems have no alternative models. There should be no choice but to accept the already formulated unbelievable conclusion. The effect of believability on one-model problems is, therefore, better explained as arising at a post-reasoning stage of filtering. In everyday reasoning people can expect to arrive, in most cases, at believable conclusions. They should, therefore, be wary of unbelievable conclusions. As we discussed in the introduction, there are two possible responses to an unbelievable conclusion that do not require further processing of mental models. The first is to decide that, since the only half-way reasonable conclusion is unbelievable, there must be no valid conclusion. The second is to decide that some minor error has produced a conclusion related to the true one, perhaps by a "superficial" change such as deleting *not* or changing *all* to *none*. In fact, the most common incorrect response for the one-model problems with unbelievable correct conclusions was that there is no valid conclusion.

The data from the multiple-model problems, especially the indeterminate ones (see below), suggest that filtering is not the only locus of the effect of believability. What is more probable is that subjects respond that nothing follows from unbelievable one-model problems because the valid unbelievable conclusion is filtered out after the search for an alternative conclusion has failed. The reason why the search fails is that there is no other model of the premises.

Determinate multiple-model problems

For the determinate multiple-model problems there were no effects of believability in any of the experiments. There are two distinct aspects to this lack of an effect. First, believability did not affect the number of correct responses. Second, it did not affect the number of suggested responses that were either produced (Experiments 1 and 2) or accepted (Experiment 3). We will consider these two aspects of the results in turn.

For the particular problems that we used the valid conclusion was always believable, so believability could not have a differential filtering effect on valid conclusions. If it was going to have an effect it would clearly have to be on the processing of alternative models. However, we found no evidence that subjects go on to produce the correct conclusion more often when the suggested conclusion is unbelievable. In fact, there were so few correct conclusions that there was no possibility of detecting such an effect. It appears that subjects have great difficulty in finding alternative models of the premises

for these problems, and in generating a conclusion that is true in all the alternative models. Indeed, the finding that subjects produce a high proportion of "NVC" responses to these problems suggests that they are aware that there is no straightforward representation of the premises that will yield a conclusion.

The data for the suggested conclusions (produced in Experiments 1 and 2 and evaluated in Experiment 3) are more illuminating. If subjects do not look for alternative models when they have found a plausible conclusion supported by at least one model of the premises, they should produce or accept more suggested believable conclusions than unbelievable ones. In Experiments 1 and 2, there was a slight tendency for subjects to produce more suggested conclusions in error when those conclusions were believable than when they were unbelievable, but in neither case was the difference significant. However, even when subjects *evaluated* suggested conclusions for these problems, there was still no indication of an effect of bias. Indeed, more unbelievable than believable suggested conclusions were accepted, though the difference was not significant. Overall, this pattern of results for the determinate multiple-model problems is difficult to explain on the assumption that believability affects only the process of constructing alternative models. It cannot be entirely the result of filtering either, because believability did not have any effect on the number of suggested conclusions produced or accepted.

One further aspect of the data from these problems is worth noting. The subjects did not simply respond "NVC" when a suggested conclusion, either presented (Experiment 3) or produced (Experiments 1 and 2), was unbelievable. There was no difference in the numbers of erroneous NVC responses for the believable and unbelievable problems.

Indeterminate problems

The data from the indeterminate problems are remarkably clear and consistent across the three experiments. In the first two experiments, suggested invalid conclusions were produced significantly more often when those conclusions were believable than when they were unbelievable. Furthermore, the correct conclusion (that nothing follows) was produced significantly more often when the suggested conclusion was unbelievable than when it was believable. Similar results were obtained in Experiment 3, in which subjects were given conclusions to evaluate. If the given conclusion was believable, subjects tended to accept it in error, but when it was unbelievable they usually correctly rejected it. In addition, they more often went on to produce the correct conclusion when the given conclusion was unbelievable. A complementary pattern of results was found for the indeterminate problems pre-

sented with the correct statement that there was no valid conclusion. When the suggested conclusion was believable, subjects were more likely to reject the correct conclusion and replace it with the suggested one.

These results support the idea that believability curtails the examination of alternative models. This hypothesis claims that subjects prematurely accept a believable conclusion based on one model of the premises, but search for alternative models if their initial conclusion is unbelievable. The results cannot be reconciled with the idea that the effect of believability arises *only* at the stage of filtering conclusions, for two reasons. First, the model-manipulating process, which *ex hypothesi* is not affected by believability, should produce the correct conclusion: NVC. But this conclusion cannot be rejected on the basis of beliefs about what the world is usually like. "No valid conclusion" cannot be judged unlikely on the same basis as:

None of the actresses is a woman

Second, it is not possible to produce an actual conclusion by making superficial changes, in the sense described above, to "no valid conclusion". At the very least it would be necessary to look back to the premises to supply the content for a believable conclusion to replace the allegedly unlikely NVC.

Conclusions

Taken together, the data from the one-model and indeterminate problems suggest that believability has an effect at *both* loci in the reasoning process. It can curtail the examination of alternative models and it can affect the final assessment of a putative conclusion. In the case of one-model problems, believability determines whether subjects allow a conclusion as valid. In the case of indeterminate problems, it affects how thoroughly they search for alternative models of the premises to refute conclusions. We cannot explain our results on the basis of just one locus of effects. On the one hand, the subjects produced a high proportion of erroneous NVC responses to the determinate multiple-model problems, and their responses were not affected by the believability of the suggested conclusions. Hence, the filtering of responses in terms of their believability cannot be the sole mechanism. On the other hand, the effects of believability on one-model problems cannot arise during the consideration of alternative models, since there are none. Hence, the effects of believability on the processing of alternative models cannot be the sole mechanism. Believability must exert its effects at both loci.

This conclusion leaves one question outstanding. Why are the results so clear for the indeterminate problems and not for the one- and, more espe-

cially, the determinate multiple-model problems? Why doesn't the believability of the suggested conclusions for these multiple-model problems affect whether they are accepted? One explanation could be that the believable and unbelievable conclusions are more polarised in the case of indeterminate problems. For example, the unbelievable conclusions might be less believable for the indeterminate problems than for the determinate multiple-model problems. However, we can immediately rule out this possibility, since it is not supported by our rating data.

A more viable explanation is that indeterminates are at the right level of difficulty for bias to have an effect: one-model problems are too easy and determinate multiple-model problems too difficult. In all three experiments subjects got a high proportion of one-model problems right. Even if a conclusion was unbelievable, they could usually see that it followed from the premises. The determinate multiple-model problems are so difficult that subjects may be unsure that even a plausible conclusion follows and unsure whether they have checked out all the alternative models. We will try to explain in more detail what we mean by the "right level of difficulty".

In the one-model problems we chose, and indeed in all one-model problems, the model always contains *obligatory* a's and c's (in the sense of Johnson-Laird & Bara, 1984a) that either must or must not be identified. For example, the model for the premises:

Some of the communists are golfers
All of the golfers are capitalists

is:

communist	=	golfer	=	capitalist
O communist	=	O golfer	=	capitalist
				O capitalist

If the premises are true, the link between the top communist and the top capitalist cannot be broken (unless an equivalent link is made), so an unbelievable conclusion at least as strong as:

Some of the communists are capitalists

must follow from the premises. Other aspects of the model eliminate the stronger:

All of the communists are capitalists

However, on the reasonable assumption that the obligatory links in mental models are the easiest to reason from, this model clearly signals that the premises support an unbelievable conclusion. In this sense one-model prob-

lems are too easy to produce believability effects. People can see that certain conclusions must follow from the single model of the premises, even if those conclusions are unbelievable.

Turning to the indeterminate problems, a closer look at them shows how an implausible suggested conclusion draws attention to a model in which that conclusion is untrue. For example, consider the syllogism:

All of the Frenchmen are wine drinkers
Some of the wine drinkers are gourmets

Many subjects drew from it the invalid conclusion:

Some of the Frenchmen are gourmets

However, the corresponding syllogism with an implausible suggested conclusion shows why this conclusion is incorrect. From:

All of the Frenchmen are wine drinkers
Some of the wine drinkers are Italians

it is not valid to conclude that:

Some of the Frenchmen are Italians

Most people see the reason: there can be two quite separate sets of wine drinkers: those who are French and those who are Italian. In both types of indeterminate problem used in the experiment, the implausible content pointed to the existence two distinct subsets of members of the category denoted by the middle term (e.g., wine drinkers). In the corresponding problems with believable suggested conclusions, no such effect occurs. It is not immediately apparent that the Frenchmen and gourmets could be members of distinct sets. Indeed, general knowledge (or at least knowledge of stereotypes) suggests that the sets are probably not distinct.

The reason why the indeterminate problems are not as easy as the one-model problems, and hence why believability can have a stronger effect on the conclusions drawn from them, is that there is no link between an obligatory a and an obligatory c that either must or must not be present in both models. If there were, there would be a valid conclusion. There is no single piece of structure in either model which shows that there is no valid conclusion, whereas part of the structure for a one-model problem shows that an unbelievable conclusion must be valid.

For the determinate multiple-model problems—those that we used, and indeed such problems in general—the difficulties of the one-model and indeterminate problems are compounded. On the one hand, there is no single piece of structure that shows that an unbelievable conclusion is valid. On the

other hand, an alternative to the simplest model, and hence to the conclusion it suggests, is not obvious. Consider, for example, the syllogism:

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

Johnson-Laird and Bara (1984a) suggest that simplest model is:

married person =	houseowner
O married person	O houseowner

	husband
	husband

This model suggests one of the following conclusions:

None of the married people is a husband
None of the husbands is a married person

The two alternative models are one in which some, and one in which all, of the husbands are married. Why doesn't the syllogism suggest these models that support much more believable conclusions? We believe that there are two reasons. First, as one of the models, illustrated below, shows, the construction of alternative models requires careful consideration of connections involving *optional* elements in the model.

married person =	houseowner
O married person	O houseowner

O married person	husband
O married person	husband

Indeed, the alternative models in all determinate multiple-model problems can only be produced by considering possible identities between obligatory a's and optional c's, or vice versa.

Second, the married husbands that have to be considered are not, and cannot be, houseowners, yet the two premises are sentences about houseowners. The premises focus attention on houseowners, and so it is difficult to think of a model in which the implausible conclusion is untrue, since that model can only be constructed by bringing nonhouseowners to mind. This case contrasts with that of the indeterminate syllogism in which the Frenchmen and Italians were both subsets of the wine drinkers. In that syllogism one of the premises was a sentence about wine drinkers. However, it is only for multiple-model problems in Johnson-Laird and Bara's (1984a) fourth figure (B – A, B – C) that both premises are about b's in this sense. So, although

the first reason for the difficulty of finding alternative models in determinate multiple-model problems is a general one, the second is not.

Our explanation of the effects of believability in syllogistic reasoning incorporates the idea of how readily reasoners can consider alternative models of premises. For one-model problems it is clear that there are none. For indeterminate problems, unbelievable conclusions are rejected because an alternative model in which there are two different groups of people that satisfy the middle term is relatively easy to think of, though not so easy that subjects are misled when the suggested conclusion is believable. For determinate multiple-model problems alternative models in which suggested, but invalid, conclusions are untrue are more difficult to construct, for a variety of reasons that depend on both the content of the premises and the distinction between optional and obligatory elements in mental models. The model construction process on which believability has its effect is, therefore, neither content-independent nor equally adept at dealing with obligatory and optional elements of models. If this process simply built another model (where possible) when the conclusion suggested by the current model was unbelievable, believability would have as clear an effect in determinate multiple-model problems as in indeterminate ones. More generally, the level of performance on determinate multiple-model problems in both the present study and in previous ones (e.g., Johnson-Laird & Bara, 1984a) suggests that the problem in finding alternative models in these problems has drastic effects on performance. Indeed, the comparatively high proportion of NVC responses to these problems, which is not related to the believability of the suggested conclusions, suggests that subjects are aware that they have not constructed a single correct representation of the premises, but that they have difficulty in finding alternative representations and in deciding what follows when more than one representation has been constructed.

References

- Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85, 1-21.
- Erickson, J.R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R.L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium*. Potomac, MD: Lawrence Erlbaum Associates.
- Evans, J.St.B.T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11, 295-306.
- Guyote, M.J., & Sternberg, R.J. (1981). A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology*, 13, 461-525.
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.

- Johnson-Laird, P.N., & Bara, B. (1984a). Syllogistic inference. *Cognition*, 16, 1–61.
- Johnson-Laird, P.N., & Bara, B. (1984b). Logical expertise as a source of error: A reply to Boolos. *Cognition*, 17, 183–184.
- Oakhill, J.V., & Johnson-Laird, P.N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology*, 37A, 553–569.
- Rips, L. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90, 38–71.
- Wilkins, M. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, 16, 83.

Résumé

Dans cet article on étudie à quel moment les effets des croyances jouent dans les syllogismes. On identifie trois points au cours du raisonnement où peuvent se produire ces effets: la représentation initiale des prémisses, l'examen des autres modèles possibles pour leur représentation (pour tous ces modèles toute conclusion valide doit être vraie), et le "filtrage" de la conclusion envisagée. Les effets des croyances sur le premier point sont déjà bien établis. Dans cet article on présente quatre expériences pour étudier si les croyances interviennent aux autres moments. Au cours des expériences 1 et 2, les sujets doivent tirer leurs conclusions de syllogismes qui suggèrent des conclusions crédibles ou incroyables. Dans la troisième expérience, ils évaluent les conclusions qui leur sont présentées. Les données montrent que les croyances affectent l'examen des modèles proposés et agissent comme filtres sur les conclusions envisagées. On conclut en montrant comment certains types de problèmes et certains contenus rendent plus probables certaines alternatives.