AN INVESTIGATION INTO BELIEF BIASES IN REASONING.

by

JULIE LINDA BARSTON

Department of Psychology, Plymouth Polytechnic

Thesis submitted to the Council for National Academic Awards in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

September 1986.
ACKNOWLEDGEMENTS

I am indebted to my supervisor Professor J.St.B.T.Evans for his unflagging patience and encouragement. His expert criticisms and advice have been invaluable at all stages of this research.

I would also like to thank Dr. P.Pollard and Dr. S.E.Newstead for their help and guidance, particularly during the early stages of this work.

Thanks are also due to Jane Evans for her word processing skills and her tolerance in laying out the tables and diagrams in this thesis. The results are most pleasing.

I would also like to acknowledge the financial support given by the Social Science Research Council. Without the research grant they awarded to my supervisors this project would not have been possible.

Finally, I should like to thank Chris for his understanding and encouragement especially during the last months of my work.
DECLARATIONS

1) While registered for this degree, I have not been a registered candidate for another award of the CNAA or of a University.

2) This research project was funded by a linked studentship grant to Professor J.St.B.T.Evans from the Social Science Research council (S.S.R.C. reference L80/20907/PSY) during the period from the 1st October 1980 to 30th September 1983.

3) Experiments 1 to 3 form the basis of a paper entitled, "On the conflict between logic and belief in syllogistic reasoning." It was written by J.St.B.T.Evans along with myself and P.Pollard and was published in Memory and Cognition, 1983, Volume 11, part 3, pages 295 - 306. The paper is contained in Appendix VI of this thesis.

4) A course of advanced study has been completed, in partial fulfilment of the requirements for the degree, consisting of guided reading in the area of deductive reasoning, supervised by Professor Evans; attendance at an advanced course on thinking and reasoning (B.Sc. Psychology, Year III special subject) and attendance at relevant professional conferences.
ABSTRACT

An Investigation into Belief Biases in Reasoning

By J.L. Barston

This programme of research investigates the effect of belief bias in syllogistic reasoning. Belief bias is conventionally characterised as a non-logical tendency to accept or reject deductive inferences on the basis of belief rather than logical argument. However, some theorists have argued that the effect is weak compared with that of logic and that it arises from misinterpretation of the premises or failure to accept the logical task.

Despite the adoption of controls recommended in the recent literature, Experiments 1 to 3 found consistently strong belief bias effects on the syllogistic evaluation task. However, there were equally strong effects of logic and an interaction between the two factors. Verbal protocol analysis revealed some possible misinterpretation of premises. More strikingly, however, it suggested the presence of three different modes of reasoning which were forward, backward or conclusion based and associated respectively with increasing levels of belief bias.

Belief bias was not observed in Experiments 4 and 5 which employed similar problem content on the syllogistic construction task. However, in view of findings recently published by other researchers, it appears that more salient beliefs are needed to produce the effect on this type of task.

Experiments 6 to 9 investigated the cause of the logic times belief interaction observed here and in earlier published studies: in essence, the effect of belief is stronger on invalid than valid problems. This could be due to misinterpretation of the logical concept of necessity, but extended instruction on logical interpretation failed to eliminate the effect. The findings were more consistent with a selective scrutiny model of belief bias which claims that arguments supporting unbelievable conclusions are more thoroughly analysed than those supporting believable conclusions. This model is discussed with reference to contemporary theories and findings in the psychology of reasoning.
## CONTENTS

**SECTION 1: REVIEW**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1</td>
<td>The rationality debate: General theoretical issues.</td>
<td>3</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>A selective survey of influential factors in human reasoning.</td>
<td>20</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Theories of syllogistic reasoning.</td>
<td>39</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>The belief bias effect: A characterisation and search for determining factors.</td>
<td>74</td>
</tr>
</tbody>
</table>

**SECTION 2: EXPERIMENTS**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5</td>
<td>Experiments 1 to 3. Belief bias and the syllogistic evaluation task.</td>
<td>101</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Experiments 4 and 5. Belief bias and the syllogistic construction task.</td>
<td>147</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>Experiments 6 to 9. The logic x belief interaction - the search for an explanation.</td>
<td>193</td>
</tr>
</tbody>
</table>

**SECTION 3: FINAL DISCUSSION**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 8</td>
<td>A re-examination of the belief bias effect in the light of Experiments 1 to 9. - steps towards a theoretical explanation.</td>
<td>233</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>The selective scrutiny model of belief bias. An interpretation of empirical evidence and a consideration of wider issues.</td>
<td>259</td>
</tr>
</tbody>
</table>

REFERENCES. 233

APPENDICES 1 to 6.
SECTION ONE
REVIEW

CHAPTER ONE
The rationality debate:
General theoretical issues.

CHAPTER TWO
A selective survey of influential factors in human reasoning.

CHAPTER THREE
Theories of syllogistic reasoning.

CHAPTER FOUR
The belief bias effect:
A characterisation and search for determining factors.
Chapter 1

The rationality debate: General theoretical issues

1.1 The issue of rationality 7

1.2 Reasoning without logic 12

General conclusions 19
The psychological study of syllogisms began around the turn of the century and coincided with popular interest in the evolution of Man. Not long before, Darwin had proclaimed Man's kinship with lower animals and this, understandably, cast doubt on the hitherto unquestioned belief that Man was unique by virtue of his rationality. The ensuing growth and refinement of experimental enquiry into syllogistic reasoning has documented a strong and systematic deviation from the dictates of formal logic.

A large proportion of research has focused upon an apparent distortion of reasoning due to problem content. One of the earliest studies to report this effect was Wilkins (1928). Although, in this case, performance was uncharacteristically impressive overall, subjects seemed to do particularly well when the problem content was meaningful as opposed to abstract or unfamiliar. However, this improvement was subdued for those problems on which the logical validity of the conclusion conflicted with pre-existing beliefs about its truth status. Subsequent studies have improved upon this design and provided substantial evidence suggesting some form of belief-based effect; which is conventionally described as a tendency to accept or reject conclusions according to their inherent believability, without regard to their logical status. The implications of such findings are clear. According to standard propositional logic, an argument must be assessed strictly according to its form and not
its content (see Appendix 1 for an introduction to the logic of syllogisms). If the reasoning of intelligent adults rests upon the guise in which a problem is presented we must re-evaluate the role which logic is assumed to play in deductive problem solving.

The effect of belief is seldom disputed. However, the limited scope of existing studies allows it to be characterised in widely differing terms which vary in the extent to which rationality is impugned. There have been many studies of belief bias, which focus almost exclusively on the syllogistic reasoning task, but there is little variation in content type and paradigm. Work in other areas of deductive reasoning shows the effect of problem content to be labile and hence it is not clear to what extent one may generalise from the existing belief bias studies. In Chapter 2 we examine recent research into the effects of content and paradigm and also consider other pertinent research into linguistic effects and error due to quantifier misinterpretation. Knowledge and understanding of such factors is essential to the interpretation of existing studies and aids the design of future belief bias research using syllogisms.

On the basis of evidence collected so far we can tentatively identify certain characteristic features of the belief bias effect. Gordon (1953) has noted that belief bias is relatively weak compared with atmosphere bias (Woodworth & Sells, 1935). Furthermore, Kaufman & Goldstein (1967) have shown that belief bias interacts with logic. Feather (1964; 1967) has suggested that the strength of an opinion is positively related to degree of belief bias and hence dogmatism is seen to be an associated factor.
Because of its relationship to personality factors and the manner in which initial findings were interpreted, some researchers seem to view belief bias as motivational in nature (e.g. Morgan & Morton, 1944). However, no such causal relationship can be inferred from existing studies. Other researchers (e.g. Henle & Michael, 1956; Revlin & Leirer, 1978) have attributed belief bias to the rejection of the logical task or distortion of problem premises. Although a few extreme cases of the former were reported in a cross-cultural study reported by Scribner (1975), such examples were only occasionally encountered. There is slightly more evidence for the latter explanation, but even this is somewhat dubious. In drawing attention to this possible cause Scribner, like Henle (1962), places much emphasis on protocols which seem to imply that belief bias (or what she terms 'empirical bias') acts primarily as a selector or editor of the evidence. The use of protocols provides a useful source of data, but we must interpret them with care and due consideration. We cannot presume that they reveal why or how a subject goes about solving a problem, especially when the verbalisation is retrospective. The way in which the protocol is elicited and the conditions under which it is produced are important factors. Indeed, other studies by Scribner (1975), with the Kpelle people, suggest that subjects rarely fail to attend to the premises initially; they become distorted only after a response has been given. In other words this distortion of the premises may occur as a result of a post hoc rationalisation process and not an encoding failure (c.f. Wason & Evans, 1975; Evans & Wason, 1976; Nisbett & Ross, 1980).
If we consider the current state of theoretical advancement and the implications of existing studies two important questions are raised:

(i) Are people incapable of rising above their inner feelings to assess the evidence on the basis of its logical merits? If so, should we call them irrational?

(ii) Can we explain the effect of belief by appealing to existing theoretical viewpoints or do we need to postulate alternative explanations and mechanisms of inference? If so, what should they be?

In the remainder of this chapter we consider the first of these questions and touch on the second by assessing some general theories of deductive reasoning. More specific theories of syllogistic reasoning are considered in Chapter 3. The final chapter of this review returns to specific studies of belief bias which are interpreted in the light of pertinent findings covered in the first three chapters.

1.1 The issue of rationality

The suggestion that people could be insensitive to the laws of formal logic has generated much controversy. Adams (1980) has cited historical support for the counterview, arguing that Man could never have cracked genetic codes, cured deadly diseases or conquered space if he were not capable of high level inference. Clearly Man's evolution can be interpreted in such a way as to support both views (e.g. Zajonc, 1980). Unfortunately, the way in which the rationalist argument is put leads to an equally ambiguous outcome when we come to consider the experimental
Essentially, rationalists argue that people are capable of logical reasoning, but that errors are an inevitable consequence of problem misconception; which may arise for various reasons. A leading proponent of this view is Mary Henle, who is one of the few to find no evidence of any belief bias effect (see Henle & Michael, 1956). In a celebrated publication Henle (1962) argued strongly that the answer given to every reasoning problem was the 'right' one, but that the question answered may often be the 'wrong' one i.e. not the question intended by the experimenter. People may misinterpret and omit premises or even reject the logical task. In order to characterise the reasoning process as non-logical it is necessary to assume that subjects interpret the reasoning task correctly (Smedslund, 1970). Since this cannot be assumed, the argument becomes a circular one. Viewed in this way the rationalist argument is scientifically untenable because of its inherent irrefutability.

Richter (1957) has attempted to get around this problem by suggesting that a measure of true reasoning ability may be achieved if all possible distorting factors are eliminated. The fundamental problem with this type of approach is that it may well be impossible to discover or agree upon a 'pure' reasoning task. Such a quest is clearly inductive; although such a task may be agreed upon initially, in retrospect it may be possible to reinterpret responses in the light of a newly apparent misinterpretation or previously undetected non-logical factor.

The rationality debate has grown more complex since Henle's seminal publication. Theorists have challenged the assumption
that formal logic should form the basis of a competence system. Braine (1978) has pointed out the discrepancy between natural language and formal logic and has followed the lead of Osherson (1975) and Johnson-Laird (1975) in proposing that inference schemata should take the place of the latter. The distinction between natural language and logic is an important one, which receives a great deal of attention in other areas of research (see Chapter 2). However, theories of inference schemata are problematical. Because they retain a competence/performance distinction they are difficult to test. Since inference schemata are symbolic in nature this type of explanation would need to postulate some form of content based encoding process in order to account for content effects. This not only reduces the parsimony of inference schemata theories, but also gives rise to questions concerning the mechanism responsible for encoding content and contextual variation in sentences. Indeed, as Evans (1985) has pointed out, encoding the meanings of sentences according to context would necessitate the use of inference rules. Hence, if reasoning and encoding are mutually reliant we are faced with the proverbial chicken and egg problem.

Recently, Johnson-Laird has changed his theoretical views on the nature of logical mechanisms in reasoning. In a major work devoted to the theory of mental models, Johnson-Laird (1983) has argued convincingly against theories of mental logic. In addition to the criticisms pointed out here, he draws attention to further, fundamental, pitfalls. If there is a mental logic, how did it come to be there? There seems to be no developmental source and it is difficult to see how it could be inborn.
Furthermore, logic allows a variety of valid conclusions for any given problem; it favours none in particular, yet people tend to produce one specific type of conclusion and seem to think that other equally valid conclusions are irrelevant and even absurd. Additionally, no single logic can account for the variety of inferences which people make. Thus, even if we 'disregarded' questions about the form of mental logic and the testability of the theory, such an explanation would still be both explanatorily and descriptively inadequate.

Other critics of logic based competence systems have offered less specific alternatives. Cohen (1981) has proposed that intuition should be the arbiter of our decisions. Any response that has the backing of consensus should be classed as valid and hence rational. Only if, under 'ideal' conditions, the subject changes his mind about his response can we then categorise it as an error. Paradoxically, once the subject recognises his error, his rationality remains intact!

The influence of problem content is difficult to explain in rationalist terms. Cohen has suggested that subjects are frequently duped by the experimenter on such occasions. They fall prey to a cleverly concocted "cognitive illusion" which acts to trick the subject into logically inappropriate analysis. Wason (1981) has noted that the world is full of cognitive illusions. Since visual illusions can add to our knowledge of perception we may yet learn something useful from the tricks we play on subjects.

Like Cohen, Kyburg (1983) rejects formal logic, but argues along different lines. In attempting to defend the rationality of
certain systematic response tendencies (which many deem to be erroneous), Kyburg claims that people, quite rightly, evaluate conclusions according to their practical certainty. In other words, according to the likelihood of their being correct. He argues that certain tendencies, such as confirmation bias (e.g. Mynatt, Doherty & Tweney, 1977; 1978) are quite reasonable and rational. Hence, belief bias which could be seen as a type of confirmation bias, would be classed as rational in Kyburg's terms. We seem to have turned full circle now. Belief bias seems to be a threat to and yet an instance of rationality! Yet, however we define rationality it does not explain precisely how or why belief bias occurs.

In summary, we can say that the existence of errors is no great problem for the rationalist position; it is the systematic nature of such responses which presents the major obstacle. A variety of explanations has been put forward to account for these response tendencies. Theories based on formal or natural logic are unsatisfactory. They are largely untestable and both descriptively and explanatorily inadequate. A chief stumbling block is the effect of problem content. Some argue that such effects are merely the result of conjuring tricks and others imply that they are, in fact, an instance of rationality. Notwithstanding such arguments we are still left with the fundamental question: if logic does not provide the mechanism of inference - what does?
1.2 Reasoning without logic

In rejecting logic, we must formulate theories which account for everything that logic can—and more. The temptation is to propose a system which has as few constraints as possible. However, we see from Newell's theory that this is not the answer.

Newell's (1981) problem space hypothesis is proposed as a general framework covering a variety of cognitive functions, not just reasoning. Based on the original formulation of Newell & Simon (1972) the problem space can take various forms, ranging from one based on first order predicate logic to natural language. It is a symbolic network of information which is traversed by constructing paths linking related informational states or nodes. Each individual can construct his own unique problem space, which may be altered within as well as between tasks. The inappropriate use of heuristics and the limitations of memory constitute the major sources of error. As problem spaces are various and may be transient it is difficult to predict exactly what type of response will occur. Sensitive to the testability issue, Newell states five general predictions made by the theory. However, as Rips (1983a) has pointed out, none of these predictions is specific enough. Some are not peculiar to the problem space hypothesis and others rely on consensus (amongst theorists and subjects) on matters of task structure. Furthermore, Newell fails to explain how the decision-making mechanism (the architecture) operates, and on what basis it could be formed. The problem space hypothesis thus suffers from descriptive and explanatory inadequacy as well as predictive impotency. These failings underline the need for a theory which
postulates a specific and invariant mechanism whose outcome is limited by identifiable problem characteristics. The mental models theory (Johnson-Laird, 1983) has taken a major step in this direction.

Johnson-Laird argues that human inference is based on a simple system which transforms input into working models, which in turn are manipulated in order to discover their implications. The mental model is a symbolic representation which duplicates the relation structure of that which it represents. This is fundamental to the mental models theory because if it is possible to capture the meaning of a proposition in terms of a mental model then the logical properties of the relation expressed will, as a consequence, be inbuilt. Thus, we can reason according to logic, but without formal logical rules. In explicit, as opposed to implicit inference, we need to establish the validity of a conclusion by undertaking an exhaustive search for contradictory models of the information given. Because of working memory limitations and possibly a misunderstanding of necessity, subjects rarely carry out an exhaustive search. As a result of this limitation, the order in which models are built is of prime importance. The order of model construction may be governed by Gricean principles (e.g. Grice, 1975), which suggest that the most obvious models are constructed first. It is, however, difficult to define the conditions which would influence the relative salience of models. It is tempting to appeal to Pollard’s (1982) availability formulation to clarify this point. However, as we shall see, this too has problems.

Work with classical syllogisms in particular has produced
supportive evidence for the mental models theory. Certain syllogisms may be solved correctly without an exhaustive search for contradictory models. Others necessitate an exhaustive search to ensure correct solution. As the mental models theory predicts, the latter type of syllogism is, in fact, associated with the greatest frequency of errors (see Chapter 3.2). However, the mental models theory does have weaknesses. We do not know the nature of the mechanism which judges the consistency of models, and we cannot state precisely which models will be considered first and when. The latter type of problem is not peculiar to the mental models theory, but it nevertheless reduces its predictive power. This shortcoming is receiving theoretical attention from other quarters. Theorists are currently attempting to discover which aspects of information receive preferential treatment and when. Evans (1984) argues that until we know exactly what people are reasoning about, it is premature to ask how they are doing it. Pollard (1982) takes the extreme view that people do not actually reason in any formal sense; they simply give the response which springs to mind first.

Approaching the problem from an angle which is complementary to both mental models and availability, Evans (1983a) has argued that the majority of responses to reasoning tasks can be attributed to performance factors and not to any putative competence. Because of their systematic nature, errors can be directly related to semantic, linguistic and perceptual cues rather than problem structure. Furthermore, correct responses (as judged by normative systems) also frequently follow this pattern. On the basis of this characterisation Evans has claimed that
reasoning is directed by a mental set, based on pre-existing experience, which channels thought; inducing selective attention to certain aspects of the reasoning problem. The theoretical task being to identify a) the process underlying selective attention and b) the process responsible for analysing the information focused upon. Evans (1984) has characterised both processes, labelling the former 'heuristic' and the latter 'analytic'. In this, two stage, theory of general reasoning the first, heuristic, stage is rapid and preattentive. It is not, therefore, available to introspection. Heuristic processes function to select those aspects of the problem deemed, by the subject, to be relevant to its solution. Information which is not selected at this stage is judged to be irrelevant and is processed no further. The proposed operation of both processes can be illustrated with reference to a variety of reasoning tasks, including Wason's (1966) selection task and syllogistic inference. (See Evans (1982a Chapter 9) for a review of the selection task and Evans (1983a). This also refers to studies of statistical inference, which are not directly considered here.) Neither the heuristic nor the analytic process has yet been fully characterised. However, Evans has pointed out that the latter need not be based on abstract logical or natural inference rules. It could be based, instead, on direct or analogical usage of everyday inference based on past experience. Both the heuristic and analytic processes are judged to be influenced by prior experience - but it is difficult to say exactly when. Clues as to how past experience may operate at the analytic level may be obtained by analysis of verbalisations. Evans has argued that
when subjects provide a reasoned description of how they have solved a problem, they are, in effect, using their analytical processes to work out the problem of how they came to produce a given response. This may or may not coincide with the actual process of inference underlying the original response. A similar interpretation of protocols was earlier developed by Wason & Evans (1975) as part of their Dual Process theory. The heuristic/analytic theory is closely related to the Dual Process theory although there are important differences (see Evans (1984) for a comparison).

Since the heuristic/analytic theory allows the omission and misinterpretation of logically pertinent information as a result of heuristic channelling and idiosyncratic analysis of information, it may be subject to testability limitations at this point. In some respects Evans' theory resembles Newell's (1981) problem space approach and hence its predictive power is questionable. Bearing in mind that the heuristic/analytic theory is proposed as a general framework, its chief contribution at this stage will be the inspiration of further, directed, research. It would seem that this is Evans' chief objective.

Pollard (1982) has put forward an availability theory of reasoning which is allied to the heuristic stage of Evans' theoretical framework. Notably however, Pollard does not envisage any process akin to analytical processing. Based on the work of Tversky & Kahneman (1973), this theory claims that reasoning responses are mediated by available cues, that is those which are most easily brought to mind or retrieved. Hence, the response is seen to be "a function of the cue and not a process". Dominant
response patterns on the 2,4,6 problem (e.g. Wason, 1971), the abstract THOG problem (Wason & Brooks, 1976) and the abstract selection task are convincingly attributed to the availability or salience of cues, resulting from the way a problem is presented or how a question is posed. Pollard's explanation of some thematic effects is less persuasive, however. The chief problem here concerns the failure to distinguish adequately between those cases in which the most 'available' association will dominate and when it will not. Recognising this difficulty, Pollard argues that the scenario in which a problem is placed may alter the relative strengths of available cues. Once we introduce arguments of this nature it is difficult to envisage findings which could not be explained by availability. Perhaps the findings of Revlin & Leirer (1978) provide an exception, however. In this study of belief bias in syllogistic reasoning there was no apparent scenario, yet subjects often avoided the most obvious conclusions. When the valid conclusion presented for selection was unbelievable, subjects did not opt to select an alternative, believable, conclusion; instead they chose a neutral (i.e. non-propositional) conclusion. They, therefore, failed to choose any of the most obviously available (believable, logical or atmosphere based) conclusions. It is difficult to appreciate how this response pattern could be interpreted in terms of availability. Clearly, as various theorists have been quick to point out, availability is a necessary but not a sufficient prerequisite for a given response.

Wason (1983) is another theorist who argues that reasoning is deep seated in experience. Wason supports the view that
schemata or knowledge structures borrowed from long term memory can 'short-circuit' information analysis. This view is entirely consistent with those of Newell (1981), Johnson-Laird (1983), Evans (1984) and Pollard (1982). However, the central problem of when certain schemata will be utilised is not directly addressed by Wason. If, as Nisbett & Ross (1980) claim, acute or transient availability of schemata is probably an important determinant of their use, the idea of schemata utilisation will inevitably suffer from the limitations of the more parsimonious availability theory.

Of the five alternatives to mental logic considered in this section, none constitutes a complete account of human reasoning. Some have more crucial weaknesses than others. The problem space hypothesis, as it stands, cannot be considered a serious contender. It predicts almost anything and everything, which simply serves to complicate the issue. Availability appears to be the most parsimonious concept explored, but this explanation has its problems. It has little predictive power and although it is reasonably adequate in explanatory terms this is of little consequence if it is descriptively inadequate. The three remaining: mental models, heuristic/analytic and schematic theories seem to complement each other. The mental models theory provides a novel approach to logical analysis and successfully accounts for both 'correct' and 'incorrect' responses. However, it cannot yet predict exactly when and how the order of model construction will be influenced. This is crucial if the theory is to be completely descriptively adequate and generalisable. The schemata theory provides an interesting pointer, but it is not...
adequately developed at this stage. The heuristic/analytic theory is also in its infancy and will benefit from future empirical evidence. In particular we need to establish strict criteria to determine what type of information will receive selective processing and when.

**General conclusions**

We know that people make systematic errors in reasoning, and various explanations have been put forward. The traditional, rationalist interpretation is unsatisfactory because it is largely untestable and both explanatorily and descriptively inadequate. Nouveau rationalists have rejected formal logic as the basis of thought, yet the alternatives provided either do not account for content effects or deny their irrationality. The difficulty lies in providing an acceptable alternative to rationalism. Theories of reasoning without logic take various forms. A major problem with most is their predictive power. Try as we may, it is still not clear exactly when certain non-logical effects will occur. A chief stumbling block seems to be the effect of problem content. However, there has been one major step forward. We can explain how to manipulate information logically without recourse to formal rules of inference. The next step is to discover exactly what determines which aspects of information receive preferential treatment. In the next chapter we examine what may be some clues.
Chapter 2

A selective survey of influential factors in human reasoning

2.1 Research into the differences between logical and natural language usage 22
2.2 The interpretation of quantifiers 25
2.3 Thematic content 30
2.4 Effects of problem presentation 34
Conclusions 36
In order to evaluate the effect of beliefs on reasoning performance we must be conversant with the various effects which may occur in addition to or even interact with the major focus of study. Although the problem of principal interest in this thesis is the categorical syllogism, it would be difficult and injudicious to isolate this area of study from others. It would be beyond the scope of this thesis to attempt an exhaustive survey of the many areas of human reasoning research. This chapter aims simply to acknowledge and note the relevance of findings in related areas of research.

A variety of influential factors have been identified. For our purposes they may be broadly categorised as linguistic, content and presentation factors. With reference to linguistic factors theorists have long recognised that the implicit communicational codes of everyday discourse do not generally coincide with the rules of formal logic. In other words, when we communicate with each other we make certain assumptions (see Grice, 1975). If such assumptions are carried over into our analysis of logical problems, specific errors may ensue. Furthermore, although formal logic does not accommodate problem content this factor is known to exert a strong, but inconsistent, effect on performance. This may be due to the idiosyncratic interpretation of the premises, or it may be due to the adoption of external yardsticks by which conclusions are judged or decisions made. Both types of explanation imply that subjects
frequently rely on general knowledge or beliefs when familiar, real-world, terms are incorporated into problems. Sometimes performance on thematic problems is astonishingly good and some theorists have argued that this reflects the disambiguation of problem premises or even the focusing of attention on to logically relevant aspects of the problem. At other times logical performance is alarmingly poor and sometimes simply mediocre. The effect of thematic materials is indeed paradoxical. A third influential factor is that of problem presentation. Differing paradigms, using logically equivalent problems, often yield differential response patterns. The interpretation of task demands, the interpretation of propositions and the distribution of attention have all been linked to differences in problem presentation.

2.1 Research into the differences between logical and natural language usage

That some interpretations of propositions may appear to be more natural than others has been indicated in several areas of research. Propositional reasoning experiments have suggested that the conditional if $p$ then $q$ can imply causal or temporal relationships. The propositions $p$ only if $q$ and if $p$ then $q$ lead to clear cut interpretational differences which seem to imply that the term modified by if is usually the term to occur first in a temporal sequence (Evans, 1977a; Evans & Newstead, 1977; Braine, 1978; Evans & Beck, 1981). An important implication of this finding is that even abstract propositions carry with them
logically irrelevant a priori assumptions. Using thematic materials, Geis & Zwicky (1971) have presented illuminating examples and explanations of how logical propositions can resemble threats, commands and promises, appearing to invite specific inferences which are not necessarily logically implied. For example, the statement, *If you disturb me tonight, I will not let you go to the movies tomorrow*, seems to imply that if the speaker is not disturbed by the listener, the latter will be allowed to go to the movies, although logically this is not necessarily the case. According to propositional logic, even if the listener behaves himself and does not disturb the speaker, he may well be disappointed, since the reward of going to the movies may not be forthcoming.

With specific reference to categorical syllogisms, Begg & Harris (1982) have presented a communicational account of propositional interpretation, and argue that many errors in logical reasoning tasks can be attributed to the interpretation of propositions according to the maxims of everyday discourse. A frequent and fundamental error being the violation of the principle of minimal commitment (see also Braine, 1978). Stated simply this principle dictates that no single interpretation of a statement may be given preference over any other. Clearly, examples are evident in the work of Geis & Zwicky, where the invited inference is one of several possibilities. In explaining the departure from minimal commitment, Begg & Harris appeal to the principles of 'Completeness' and 'Asymmetry'. Fundamental to the application of such principles is the notion that syllogistic premises and conclusions are interpreted as attempts to
communicate. The interpretation is made in the light of an assumed commitment to a social contract, rather than to logic. Basically, it is assumed that the hearer has certain needs which the speaker must bear in mind in conveying information (Grice, 1975). The principle of asymmetry is based on the observation that the order of terms within a proposition has extra logical implications. As we have already seen, the order of terms in some conditional statements seems to imply differing temporal sequences. Similarly, it has been claimed that the subject term of a syllogistic premise is seen to be the more salient or general of the two terms mentioned. Logically, one is not entitled to make this assumption.

That the order of terms within a proposition should influence its interpretation is a claim that is directly opposed by the Conversion theory of syllogistic reasoning (Chapman & Chapman, 1959), although later versions of this theory (e.g. Revlin & Leirer, 1978) could be seen to accommodate such claims, but with reference to thematic materials only. Theories of conversion claim that subjects frequently reverse the ordering of terms within a proposition, so that either a symmetrical or an opposite relation is actually encoded. Clearly, according to this hypothesis, the ordering of terms has, in itself, little or no importance, or else conversion would not seem to be a likely practice.

If we look at the interpretation of syllogistic premises as pairs of statements rather than as individual propositions, we see once again that the ordering of terms within the pair as a whole, (i.e. its 'figure' - see Appendix I) can exert a strong
influence on the conclusion produced (e.g. Johnson-Laird & Steedman, 1978) and on the difficulty of individual syllogisms (e.g. Frase, 1968). The so-called 'figural effect' in syllogistic reasoning could be related to the principle of asymmetry although it has also been attributed to other factors such as the nature of particular representational formats (Johnson-Laird & Steedman, 1978) and a certain characteristic (first-in-first-out) feature of working memory (Johnson-Laird, 1983).

The second communicational principle referred to by Begg & Harris, that of Completeness, is based on the assumption that a speaker will provide pertinent and full information, expressing it in the most useful form, given the listener's particular needs. One example of the use of completeness is that, when drawing conclusions, subjects seem to avoid restatement of the premises, redundancy and statement of the trivial, (noted by Johnson-Laird, 1983). This behaviour does not conform to any logical constraint, yet most people would think any other type of conclusion somewhat absurd. Another example of how completeness works may be in the interpretation (or misinterpretation) of certain quantifiers. The question of quantifier interpretation is particularly important in syllogistic reasoning research. In studying the effects of belief bias on such tasks we must be able to understand and identify possible effects of quantifier misinterpretation in order to distinguish them from the principal focus of study. The following section considers how such effects may manifest themselves.

2.2 The interpretation of quantifiers

A popular method of investigating quantifier interpretation
is to ask subjects to select or produce Venn or Euler diagrams to represent a given quantifier. Although results seem to vary depending on whether subjects are asked to generate (e.g. Erickson, 1974) or select diagrams (e.g. Johnson-Laird, 1970; Neimark & Chapman, 1975), the majority of subjects interpret universal quantifiers such as No and All according to their logical meaning. Particular quantifiers such as some and some are not seem, however, to be given a more restricted interpretation than that allowed by standard logic. Newstead & Griggs (1983) have pointed out a confounding factor in this technique of investigation because different quantifiers are associated with differing numbers of appropriate set diagrams. Quantifiers with the least applicable number of diagrams are usually the easiest to interpret. Such a relationship is unavoidable and reduces the usefulness of this method of enquiry. However, it may suggest that apparent interpretational errors are in fact due to memory limitations, which tends to support the mental models theory discussed in the previous chapter (section 1.3).

An alternative method of ascertaining the interpretation of quantifiers was used by Sternberg & Turner (1981) and also Fisher (1981). In these experiments subjects were asked to deduce the truth of one quantified statement from another. In Sternberg & Turner's experiment, overall accuracy was impressively high (90%). Unfortunately, responses were not broken down for each quantifier, so little detailed information can be gathered from this study. Interestingly, however, when subjects were asked to judge the validity of a complete syllogism the accuracy level dropped to between 70% and 80%. Clearly, as analyses indicated,
task differences exerted a significant effect. This suggests that
the majority of errors arise at premise combination or later
stages, rather than at the initial interpretational stage of
reasoning. This runs counter to Henle's (1962) hypothesis.
Nevertheless, it must be acknowledged that performance here
remains impressive.

Fisher (1981) failed to substantiate Sternberg & Turner's
general findings and subjects showed some tendency to convert
logically unconvertible statements. (In terms of standard
propositional logic the only legally convertible quantifiers are
the universal negative (No) and particular affirmative (Some) -
see Appendix I(iv)). With reference to the illegal conversion of
All , Bucci (1978) has argued that subjects make exclusive use of
a 'structural-neutral' or set equivalence interpretation. By
using 'broad' and 'narrow' predicate items, whose real world
interpretations are sub/superset and set equivalence
respectively, she claims a facilitatory effect when premise
interpretation is thus clarified. However, the results obtained
are not as clear cut as she predicts.

Newstead & Griggs (1983) have criticised Fisher (1981) for
obscuring the conversion tendency by employing limited response
categories of 'possibly true' and 'necessarily false'. Varying
the range of response categories, they asked subjects to judge
the truth of one quantifier with reference to another. Using
'True and False' compared with 'True, False and Maybe' response
alternatives, they found no differences in response patterns.
Performance on contrary and contradictory statements approached
100%. Performance on indeterminate inferences was remarkably low,
however. For example, where \textit{Some} or \textit{Some are not} leaves the truth status of the other indeterminate, 75\% of subjects believed that one implied the truth of the other. Begg & Harris (1982, Experiment 2) did not replicate this finding; subjects were usually highly accurate, classifying such implications by circling the question mark, meaning indeterminate. Could these differential results be linked to the inclusion of instructions on the interpretation of \textit{some} in the latter experiment?

Following up this line of enquiry in a second experiment, Newstead & Griggs explained the meaning of 'some' and 'some/not'. Although this led to improvement in the interpretation of certain quantifiers, paradoxically the same instructions led to an increase in the erroneous interpretation of other quantifiers. Subjects inferred that \textit{All} and \textit{No} could be inferred from 'Some' and 'Some/not' respectively. This sort of error is consistent with Dickstein's contention that subjects do not interpret logical necessity according to its strict logical meaning. They do not seem to realise that, logically, indeterminate relations must be classified as invalid. Instead, subjects often class them as valid. Hence, in experiments where subjects are forced to choose between a 'True/False' response (as in this Newstead & Griggs experiment) it is highly likely that indeterminate relations will be classed as 'True'. Thus, 'Some' is consistent with 'All', and 'Some/not' would imply 'No'. Bearing this in mind, it is important to note that Sternberg & Turner explained the meaning of logical necessity, to some extent, as well as the meaning of 'Some'. They allowed three response categories: definitely, possibly and never true. Could this be why Sternberg
& Turner observed such impressive performance in their experiment?

Clearly, from the experiments outlined here, it must be concluded that consistent findings are hard to come by in this area. Methodological differences could be a major cause, although the sparse detail given in some reports tends to cloud this point.

The suggestion that content and/or context may influence the interpretation of individual propositions has already been discussed with reference to conditional reasoning (e.g. Geis & Zwicky (1971); see also Staudenmayer (1975), Fillenbaum (1975; 1976)). With particular reference to quantifiers, Stenning (1977) and Newstead & Griggs (1984) have argued that their precise meaning may vary according to the context (and content) in which they are embedded. Stenning's work emphasises two recurring findings. Firstly, the interpretation of quantifiers is not entirely governed by logical constraints, and secondly the interpretation of one, given, quantifier is not always consistent (like the problematical "If", the quantifier does not seem to be "a creature of a constant hue"). If this is indeed the case, then results from one paradigm may not be generalisable to other areas of research.

At the present point in time it is difficult to say exactly how people interpret quantifiers. They frequently make mistakes—perhaps because of the way we put the task to them. Even when we explain how to interpret the more confusing quantifiers, this seems to lead to greater confusion in interpreting the results obtained. Despite these difficulties it is not impossible to gain
useful information when using syllogisms as a tool in investigating effects such as belief bias. From the evidence discussed in this section we are aware of the kinds of mistake which can be made in quantifier interpretation; it is up to us to recognise and control for any such effects.

2.3 Thematic content

It is fair to say that the role of problem content has probably instigated the vast majority of theoretical interest and debate for more than a decade. On the one hand are those like Bucci (1978) and Revlin & Leirer (1973), who argue that any such effect is due to altered premise interpretation. Akin to this notion is the idea that some problem contents may clarify problem structure (Newstead & Griggs, 1982). On the other hand are those, like Pollard (1982) and Griggs & Cox (1982), who claim that facilitatory thematic effects are due to subjects reading off their conclusions or responses from memory and that no formal analysis or process need be postulated in order to explain responses. In between these two camps lie theorists like Evans (1983a, 1984) and Johnson-Laird (1983) who claim that some thematic contents alter the judgement of relevance or preference given to information for inferential analysis. The following section attempts to synthesise the findings so far.

Since Wilkins (1928) revealed an apparent facilitatory effect of non-controversial thematic material on syllogistic reasoning, a host of other experiments have demonstrated similar facilitatory effects, focusing mainly on Wason's selection task (e.g. Wason & Shapiro, 1971; Johnson-Laird, Legrenzi & Legrenzi, 1972). Other studies have replicated and sought to isolate the
influential thematic elements of the task (Bracewell & Hidi, 1974; Gilhooly & Falconer, 1974). More recent studies have attempted to investigate the generality of the effect, using a variety of problem contents. The majority of such studies failed to establish any facilitatory effect and, intriguingly, failed also to replicate the original findings reported by Wason & Shapiro and Johnson-Laird et al (e.g. Manktelow & Evans, 1979; Griggs & Cox, 1982; Reich & Ruth, 1982; Yachanin & Tweney, 1982). Only Pollard (1981) and Golding (1981b) seem to have been able to revive the original findings.

There are several detailed reviews of this area, so, to avoid unnecessary repetition, only the main issues will be covered here. For further reading, Evans (1982a) and Griggs (1983) provide up-to-date accounts.

Throughout the literature, three observations can be made. Firstly, facilitatory effects are content specific, that is to say they seem to occur on certain types of thematic content - but not others. Secondly, those studies which report an effect have, so far, failed to demonstrate any transfer effect, characterised by improvement on a subsequent abstract version of the task. Finally, as other researchers have noted (van Duyne, 1974,1976; Griggs & Cox, 1982; Pollard, 1982), a higher level of correct responses may be obtained by inducing an additional 'detective set' or presenting the problem in a plausible context.

Let us first consider the specific conditions under which an effect has been obtained. The highest reported level of facilitation was achieved by Johnson-Laird et al, using a task in which subjects pretended to be postal sorters. The rule given
concerned the stamp value on an envelope and whether or not that envelope was sealed. It has been pointed out that a rule relating stamp value and sealed or unsealed envelopes was once enforced by the British Post Office some years ago. It has, therefore, been suggested that this experiment does not demonstrate facilitated reasoning, but instead indicates that some form of memory cueing effect prompts the correct answer. Indeed, experiments by Griggs & Cox (1982) and Golding (1981), which manipulated past experience with materials, tend to support this contention. This would explain why the arbitrary pairing of food and drink terms used by Manktelow & Evans and the various other unrelated terms used by Reich & Ruth and Yachanin & Tweney were unsuccessful in producing facilitation. There is one problem with this interpretation, however. It cannot explain why Wason & Shapiro's arbitrary pairing of towns and transport produced a facilitatory effect. Admittedly this finding is unreliable, but its fickle nature may well be a clue as to its cause. Pollard (1982) suggests that it may not be the towns and transport content in itself which creates the effect; it may also be necessary to add an element of doubt (in the subject's mind) about the truth of the claim made. If the claim is presented as a rigid, specific, rule or as a boastful claim, subjects may feel inclined to check up on the claimant. It therefore follows that both the way in which the rule is presented and possibly the person who presents it are important contributory factors. Certainly, the presentation of Manktelow & Evans' food and drinks problem in the context of a plausible diet plan by Pollard & Gubbins (1982) showed that a previously ineffective thematic content may produce
slight facilitation when presented in a different manner (see also Griggs & Newstead, 1982, who demonstrate a similar effect on Wason's, 1977, THOG problem).

In characterising the effect of thematic materials we must consider the suggestion that the 'detective set' must be invoked in order to achieve facilitation; certainly all the successful attempts at facilitation could be thus explained. The fact that there are no transfer effects seems to rule out any suggestion that the facilitation is due to enlightening the subject as to the logical requirements of the task. The fact that 'set' or context may aid facilitation may be an important clue. It is difficult to see how this could aid problem representation (as rationalists would argue), however, it may well alter the subject's approach to the problem. If the subject suspects that the rule (or claim made) is, or could be, false, whilst not necessarily appreciating the need to falsify, he may make a thorough check of the alternatives. This suggests that under abstract conditions subjects may analyse problems in a superficial, or at least incomplete, manner. The finding that 'matching bias' (Evans & Lynch, 1973), an apparently superficial matching strategy, appears only when stimulus material is low in thematic content or meaningfulness (Reich & Ruth, 1982) serves to illustrate this point. Interestingly, even if no facilitatory effect is observed, a thematic effect of some sort may still occur as long as the problem is meaningful. Note, though, that facilitation seems only to occur if subjects suspect that alternatives to the given rule or claim are equally possible or even more likely.
2.4 Effects of problem presentation

Associated with effects of prior knowledge are those effects which appear to evolve from the subject's perception of task demands. In various areas of cognition, the subject has become known as an active perceiver who tries to make sense of the stimuli presented (e.g. Neisser, 1967). Deductive reasoning research provides many instances in which different problem presentation formats have led to differential response profiles for logically identical problems.

Using an abstract conditional reasoning task, Taplin & Staudenmayer (1973) noted that responses to a conditional reasoning problem differed according to the number of response alternatives allowed. When subjects were forced to make a binary true/false decision, many subjects responded as if a biconditional truth function was being employed. In a second experiment subjects were allowed to use an additional 'sometimes' response category. Under such conditions some subjects appeared to be using the conditional interpretation of the rule given. Taplin & Staudenmayer suggest that the differing response patterns may reflect the expectation that all the response categories offered should be made use of. Evans (1982 p.136) has criticised the experimental methodology used here and argues that the restricted response alternatives given in the first experiment may have forced subjects, in some instances, to choose between two erroneous responses. In some cases the indeterminate ('sometimes') category was logically necessary. Evans would, therefore, argue that the differences in response patterns across experiments could equally reflect the fact that some subjects
were availing themselves of the opportunity to make a logically necessary (indeterminate) choice when available.

Related to Taplin & Staudenmayer's suggestion is Revis' (1975) claim that subjects endorse few non-propositional conclusions in classical syllogistic reasoning tasks because they do not expect the majority of problems given to be indeterminate (which, according to Revis, they usually are). This (non-logical) expectation acts to influence premise interpretation and encourages the use of illogical strategies. This makes people appear to be less rational than they actually are. Dickstein (1981) has refuted this suggestion empirically and proffers his own explanation for the apparent bias against the non-propositional conclusion.

Other research based on classical syllogisms has shown that certain non-logical response tendencies may be due to the way in which a problem is presented. For example, Mazzocco, Legrenzi & Roncato (1974) have shown that atmosphere bias, discussed in Chapter 3, disappears when the syllogistic task is presented in an unconventional way. Ekberg & Lopes (1980) have also noted that when a greater variety of quantified conclusions is allowed, atmosphere bias fades out of the picture.

Studies in other areas of research have varied response alternatives and found differential response patterns. In the Wason selection task, Johnson-Laird & Wason (1970b), using an abstract task, and Lunzer, Harrison & Davey (1972), using a thematic task, have both indicated that a facilitatory effect may be achieved by reducing the cards to be turned over to those symbolising the consequent case (i.e. \( q \) and \( \bar{q} \)).
Overall, it seems unwise to relate performance on one specific task to pervasive underlying processes when, as we have seen, there is little consistency over logically equivalent problems of differing presentation formats. Linguistic, content and presentation effects can be manifest in different ways. Sometimes there is a high level of logical performance and at other times it is extremely low. It has been argued (Johnson-Laird & Steedman, 1978) that some paradigms do not even require the subject to reason at all. This does not imply that the responses which are produced under such circumstances are any less worthy of investigation; it does, however, emphasise that we cannot take for granted the generality or centrality of any given effect, and above all we must beware of constructing or being misled by spurious effects.

Conclusions

This chapter has focused on three influential factors, which have until recently received limited empirical attention. In the first section we saw that even when problems are expressed in abstract terms they may still be more meaningful than the experimenter originally intended. People 'go beyond the information given', giving certain interpretations precedence over others and making logically unwarranted assumptions about the meaning of the statement used. Many of the findings covered were linked to Begg & Harris's communicational account of reasoning. A central theme was that subjects tend to interpret their task in terms of an assumed social contract. A priori knowledge about the use of language tended to colour the types of inference made. In the second section we considered probably the
most obvious example of the intrusion of a priori knowledge, namely content effects. In particular, facilitatory effects were examined. Specifically, it was suggested that improvement occurred selectively on problems which evoked a 'detective set', and the notion of 'memory cueing' was shown to have restricted applicability. The detective set idea seems increasingly appropriate in the light of the discussion of certain linguistic effects, which also seemed reliant on following up 'lines of enquiry' prompted by everyday experience. It may be tentatively suggested that facilitation is due to deeper processing of certain elements of the problem, as if thought could be focused more finely under certain conditions. The idea of focusing attention was discussed with reference to the heuristic/analytic theory of Evans (1984) and it is interesting to note that attentional factors might explain some of the paradigm-specific effects considered in the final section. When logically equivalent problems are presented in different formats, we may find that our processing system finds some tasks more easy to handle than others, possibly because certain relevant (and in other cases irrelevant) features of the task are made more salient in differing presentation modes. Hence, as we saw earlier, certain 'lines of enquiry' may be prompted purely because the problem is presented in a certain way.

Linguistic, content and paradigm effects can all be used to argue for and against the orthodox, rationalist interpretation of reasoning. The illustrative examples chosen were not presented to argue for any particular position, but to demonstrate the volatile nature of human reasoning (see Evans, 1982, for further
examples). It is a fundamental scientific rule that, whilst focusing on one chosen effect, the researcher must be conversant with the various influential factors which may contaminate his findings. Each of the factors discussed here arises from the varied use of different investigative tools. Unlike the biologist, the psychologist cannot disinfect his instruments; it is this fact which, in particular, renders human reasoning a peculiarly rich scientific domain.
Chapter 3

Theories of syllogistic reasoning

The atmosphere hypothesis. 43

3.1 Error in premise interpretation 44

The conversion hypothesis. 44
Revlis' (1975a) 'feature selection' and 45
'conversion' models.
The 'Play it safe' and 'Natural 49
reasoning' models.

3.2 Error in combining premises. 51

The analogical model. 51
Mental models in syllogistic reasoning. 54
Erickson's (1974, 1978) complete and random 60
combination models.
The transitive chain theory. 61

3.3 Error due to incorrect translation of derived 64
conclusions.
Dickstein's directional processing theory. 66

3.4 Theories which consider errors at each stage of 68
the reasoning process.
Newell's (1981) application of the problem 68
space hypothesis.
Fisher's (1981) three factor model. 69

Conclusions and implications for the study of belief bias. 71
A large body of research into syllogistic reasoning has shown that performance frequently departs from logical accuracy, but that errors are not random. The principal investigative aim has been to identify the factors which contribute to differences in difficulty among the various syllogisms. Logically, syllogisms may vary according to their mood and figure (see Appendix I for a reminder). Psychologically there are several additional variables which may result in differential difficulty. To begin with, let us consider the two basic variables of figure and mood.

Figural effects are seen most clearly in studies of the syllogistic construction task, where subjects are given two premises and are required to provide their own conclusion. In this case, the effect is seen as a tendency to produce P-S conclusions on Figure 1, S-P conclusions on Figure 4 and a mixture of the two types on Figures 2 and 3. The figural effect is also manifest in other ways. On evaluation and multiple choice tasks, in which subjects evaluate a given conclusion or choose one from a selection respectively, the effect appears as a specific pattern of error rates. Overall, there are relatively low error rates on Figure 1, high error rates on Figure 4 and intermediate rates of error on Figures 2 and 3. The strength of the figural effect varies according to the mood of a given syllogism (Dickstein, 1978a; Johnson-Laird & Steedman, 1978).

Effects of mood are twofold. The well documented atmosphere effect is based solely on mood. Varying combinations of specific (A,E,I and O) propositions yield consistent response patterns. However, the strength of the effect seems to depend on paradigm. It appears to be strongest on the evaluation task. It is also
stronger on valid or determinate syllogisms. Clearly mood alone cannot account for these findings. In Chapter 2 it was noted that the four propositions which form the basis of mood are associated with varying numbers of set relations. Some theorists have pointed out that difficulty can be associated with the number of set relations to be encoded for each premise (e.g. Ceraso & Provitera, 1971; Erikson, 1974). The number of premise combinations which need to be considered in order to arrive at a correct conclusion is also an important determinant of difficulty. Johnson-Laird & Steedman (1978) report that 80.4% of the responses to syllogisms not affected by an incomplete consideration of all combinations were correct, whereas only 46.5% of responses to problems that could be affected by an incomplete test were correct (see also Johnson-Laird, 1983, pages 101-105 in particular).

Evidently, mood and figure effects cannot be considered in isolation; they are inextricably linked. The combination of mood and figure leads us to the question of validity. It is difficult to assess the relative difficulty of valid and invalid syllogisms, because accuracy levels are interwoven with figure and mood effects. However, in general we may say that people perform better on valid (or determinate) than invalid (or indeterminate) syllogisms (Chapman & Chapman, 1959; Roberge, 1970; Dickstein, 1975; 1976). However, this tendency is not evident on the syllogistic construction task (Johnson-Laird & Steedman, 1978). The majority of studies report the mean accuracy level for indeterminate syllogisms to lie around 32%, but this figure is sometimes lower. For determinate syllogisms the figure
is around 51% (e.g. Roberge, 1970). Many studies use determinate syllogisms as filler items only; hence we have relatively little information about performance on such problems.

Since performance on invalid and indeterminate syllogisms is relatively poor, theories are largely concerned with explaining the types of error found on this sort of problem. Individual publications often centre around one theory in particular, and the data are often presented from the perspective of that theory alone. Hence, it is rarely possible to assess the generalisability of specific theories across different paradigms and problem contents. We know from the preceding chapters that both variables can be influential. We also know that instructional variations affect syllogistic reasoning performance (e.g. Dickstein, 1975; Simpson & Johnson, 1966), yet many studies report their instructions in sparse detail. The evaluation of theories or specific studies depends on an appreciation of the complex interplay between various factors pinpointed in Chapter 2. The most valuable contributions are those which attempt to control for the factors outlined.

All but one of the theories considered in detail follow an information processing approach. This type of problem analysis is fitting because it is convenient to divide the syllogistic reasoning process into a number of stages relating to premise interpretation or encoding, combination and response. The following review attempts to categorise theories according to the stage which is given most emphasis in explaining error. It is not possible, nor desirable, to compartmentalise each theory completely and so there is occasional overlap of studies across
categories. Ideally, this method of organisation will enable the reader to assess the relative importance of different stages in syllogistic reasoning, without becoming bombarded with the intricacies of the many and various theoretical viewpoint.

The first explanation to be considered preceded the recent popularity of the information processing approach. It stands out from its competitors by attributing errors to a non-logical bias which is based solely on the mood of syllogisms.

The 'atmosphere' hypothesis

The atmosphere hypothesis (Woodworth & Sells, 1935; Sells, 1936) claims that syllogistic reasoning is based on a non-logical set or global impression formed by the quality and quantity of the combined premises. Thus, premise analysis is on a superficial level and figure has no role to play. Specifically, the hypothesis states that: 'If at least one premise is negative, then the most frequently accepted conclusion will be negative. If at least one premise is particular then the most frequently accepted conclusion will be particular.' The overall impression may result from the combination of these two preferences where appropriate. Thus, if one premise is particular and the other negative, then a particular negative conclusion would be accepted.

Tests of the atmosphere hypothesis have taken various forms. Instructional manipulations, warning subjects not to be seduced by global impressions, have had mixed effects (Simpson & Johnson, 1966; Dickstein, 1975). However, conflicting findings may have resulted from additional subtleties included in the instructions (see Pollard, 1979a) and failure to use comparable syllogisms.
with different instructions (see Evans, 1982a, page 91).

Shortcomings of the atmosphere hypothesis are, to some extent, masked by adherence to conclusion evaluation and response selection paradigms. Since the conclusion chosen is seen to be the result of mood alone, the atmosphere hypothesis does not predict figural effects. The nature of the figural effects is most clearly evident on the conclusion construction task (see Johnson-Laird & Steedman, 1978, and Johnson-Laird & Bara, 1984). This task also provides evidence of a strong tendency to produce non-propositional conclusions (e.g. Jackson, 1982). The atmosphere hypothesis cannot account for this type of response, since there is always a propositional conclusion available which is consistent with atmosphere.

Clearly we discover new aspects of performance as we introduce novel paradigms. Indeed, it was Sells himself who suggested that atmosphere predictions may be favoured more under certain conditions than others. Whilst this plasticity presents difficulties for theorists it does, in fact, reveal important features of the reasoning system in general.

3.1 Error in premise interpretation

The conversion hypothesis

In contrast to atmosphere, the conversion hypothesis (Chapman & Chapman, 1959) presents a traditional rationalist interpretation of errors. According to this view, people make errors because they assume that all premises are convertible. Thus, the conversion of non-convertible All and Some/not premises leads to a specific error pattern, which accounts for errors on 13 of the 45 invalid syllogisms usually used. For the remaining
32, the conversion of premises would lead to a non-propositional conclusion - which is, of course, correct. To account for errors on these problems a process of probabilistic inference was proposed. In this case, people are supposed to assume that if the subject and predicate terms of a syllogism share the same middle term in the premises, and hence have something in common, they must be related to each other. If they do not share the middle term they are assumed to be unrelated. Both conversion and probabilistic inference are claimed to result from the generalisation of assumptions which normally hold in real life. Dickstein (1981) has challenged both this rationale and the idea of probabilistic inference. His findings suggested that a misunderstanding of necessity was a more likely source of error than probabilistic inference. This explanation not only accounts for why subjects accept invalid conclusions, it also implies that performance on valid syllogisms will be better. It does not, however, predict which conclusions will be drawn or selected. It only explains why they are accepted and it cannot account for the selection or production of non-propositional conclusions.

Revlis' (1975a) 'Feature selection' and 'conversion' models. Revlis has attempted to improve the theoretical status of both the atmosphere and conversion hypotheses by recasting them as more specific processing models. In terms of overall predictions, the feature selection model differs from atmosphere (as stated by Begg & Denny, 1969) in that deviations from predictions are expected as the number of operations increases (which is the case whenever extracted premise features mismatch at the composite representation stage). This means that responses
will deviate from predictions whenever the two premises mismatch on quality or quantity (see Dickstein, 1975, and Mazzocco et al, 1974, for similar formulations of atmosphere predictions). As well as assuming that subjects automatically encode the converse of premises only, Revlis argues that subjects do not expect to be given such a large proportion of invalid syllogisms, and hence have a bias against accepting the non-propositional conclusion. However, as we have noted, work by Dickstein (1976) suggests that this is not a likely reason for the acceptance of invalid conclusions.

Revlis' conversion model differs slightly from that of Chapman & Chapman. Both premises are initially encoded in their converted form. A composite representation is formed, after which the conclusion given is encoded in its converted form and compared with the composite representation. If they are the same, the conclusion is accepted; if not, the next conclusion, if there is one, is considered. After considering all the propositional conclusions offered, the subject reworks the problem, making a second 'PASS' through. On this run through, the premise interpretation is logical. If this fails to produce a premise representation which matches a propositional conclusion, rather than accept a non-propositional conclusion, the subject makes a fair guess from amongst all conclusions given. As a result, the correct, non-propositional, conclusion will be chosen for approximately 20% of such problems (when there are five alternatives from which to choose). Because of the nature of conversion, Revlis and Chapman & Chapman predict different responses for a small number of the 64 premise combinations
conventionally used. However, recent research seems to provide more support for the Chapmans' version (see Dickstein's, 1981, review of multiple choice task data). In a test of Revlis' feature selection and conversion models, Revlis claims that the latter is the more appropriate explanation of syllogistic reasoning - despite the fact that the former model can account for more of the data. The feature selection model accounted for 71.4% of responses, whereas the conversion model accounted for 61%. The stronger of the two models is rejected because its non-logical motivating principle seems to be contradicted by the fact that it gains most support when it prescribes the logically correct (in this case, valid) conclusion: It is puzzling that Revlis should view his conversion model as being the more appropriate when, if subjects are quite capable of reasoning logically, they save it as a secondary procedure and prefer to guess if the logically derived conclusion does not suit their taste. Surely, such behaviour cannot be viewed as any more rational than feature matching.

A third model, proposed by Revlis (1975a) (See Footnote 1) incorporates the feature selection process as part of the 'GUESS' mechanism in the conversion model. Whilst this increases its predictive power, it also makes it post hoc and unparsimonious.

In evaluating any theory of premise misinterpretation we need to consider a) whether the proposed misinterpretation occurs and b) whether it also operates during syllogistic reasoning. Research on quantifier interpretation provides equivocal support for a conversion explanation. When subjects draw Euler diagrams to represent quantified propositions (e.g. Erickson, 1978) they are
more likely to exhibit conversion tendencies than when they select them (e.g., Neimark & Chapman, 1975). In addition, subjects who share out points according to the applicability of set diagrams to propositions show no preference for symmetrical propositions as Chapman & Chapman would have to predict. Rare support for conversion is provided by Newstead & Griggs (1983). About one third of subjects consistently converted All and over 60% converted Some/not propositions. However, many subjects failed to convert the No and to a lesser extent the Some proposition.

As we have already seen in the previous chapter, variation in instructions, task presentation, content and context may lead to conflicting results in quantifier interpretation. The generalisability of any effect cannot therefore be guaranteed.

Some studies of complete syllogisms have attempted to block conversion or to clarify and restrict the possible premise interpretations. Agnoli (1974) and Revlin & Leirer (1978) have argued that when relationships which are not convertible in the real world are used as premises (e.g., All dogs are animals) there is an appreciable improvement in performance. Using modified premises which made the appropriate set relation explicit, Ceraso & Provitera (1971) found that performance improved markedly. Such studies emphasise that the encoding or interpretation of premises could be problematical. However, this does not provide exclusive support for the conversion explanation. Other theories (e.g., Erickson, 1974; Caplan, 1981; Ekberg & Lopes, 1979) propose alternative forms of premise misinterpretation.

Conversion theories face additional problems. If subjects
spontaneously converted all premises, performance on valid syllogisms would be considerably worse than it actually is. A crucial problem is that conversion cannot account for figure effects; indeed it does not allow them.

The 'Play it safe' and 'Natural reasoning' models.

Erickson has put forward a set theoretic model based primarily on Ceraso & Provitera’s findings. He suggests that performance is negatively related to the number of possible premise interpretations and combinations. Because of the nature of representation, this theory too has problems in explaining figural effects. We consider it fully in the next section.

Caplan’s (1981) play it safe model is closely based on Erickson’s ideas. She claims that people play it safe by encoding only the assertion made in the premise, without going beyond it. For example, *All A are B* asserts simply that, we are not told anything about the relationship of B to A. Thus the sub/superset interpretation is the safest. The equally appropriate identity interpretation is unsafe because it goes beyond the stated relationship. There are clear Gricean associations here, and the suggestion is intuitively appealing. However, there is no direct evidence to support this idea and although Caplan claims to have supported her model the test to which it is put appears suspect. Comparing her results with those of Johnson-Laird & Steedman (1978), who performed a similar experiment, we find large discrepancies. The play it safe model can explain only 23% of responses to valid and 49% of responses to invalid syllogisms in their study. The differential response patterns could be the result of thematic effects which happen to favour Caplan’s model
in her test.

An alternative view of premise misinterpretation has been put forward by Ekberg & Lopes (1979a,b). In their natural reasoning model they propose, rather like Caplan, that premises are interpreted according to ordinary or natural language usage. In this case quantifiers are seen to be fuzzy concepts. Each quantifier may describe more than one specific quantity, but will vary in its degree of appropriateness. The model is the first of its kind to propose a process of mathematical (multiplicative) inference. This is entirely plausible, given that we are dealing with quantification and that we are all taught how to multiply at school. Because there seems to be no checking procedure it appears that subjects should always respond with a propositional conclusion. This fits their own data quite well (Ekberg & Lopes, 1979b). However, a similar experiment by Johnson-Laird & Steedman (1978) clearly shows that subjects do produce non-propositional conclusions quite frequently. Other predictions are not well supported. Since the inferential process is multiplicative the model predicts that any premise pair containing a No proposition (which is represented by 0.0 or something approximating this) should lead to a No conclusion. This is not apparent in their own data nor in those of related experiments.

Ekberg & Lopes have gathered support for some of their assumptions. However, throughout their series of experiments they use a somewhat novel technique. Subjects are required to adjust the length of a graded line to represent the quantifier which results from the combination of two quantified propositions. Their reliance on this method raises doubts about the generalisability
of their findings. Perhaps the line length manipulation task actually favours fuzzy quantifier interpretation. Representing judgements by positioning an arrow along a clearly divided line may also make a number or proportion-based inference process more likely. Certainly, experiments which do not use this technique do not find the same response patterns as those of Ekberg & Lopes.

Both Erickson's set theoretic model and Ekberg & Lopes' theory of natural deduction span the gap between theories focusing on premise misinterpretation and those that concentrate on error at the combination stage. Both the aforementioned explanations place the burden of error in interpreting and combining the premises. A move away from distorted encoding explanations is embodied in the analogical model put forward by Johnson-Laird (1975) and expanded upon by Johnson-Laird & Steedman (1978). The theory of natural deduction, which we have just considered, is closely based on the analogical model. However, there are important differences as we shall see in the next section.

3.2 Error in combining premises

The analogical model

This model assumes that premise interpretation is logical. Errors related to mood arise because of a failure to consider all possible premise combinations. Figure influences the conclusion drawn because of the way in which premises are represented at the combination stage. The two premises are represented analogically by three classes containing an arbitrary number of exemplars. Arrows connect exemplars as appropriate. For example:
At the combination stage there is a bias towards linking end terms via the same middle term. This initial, biased, representation forms the foundation for the conclusion produced. Here we have a path running from a to c elements. Since it contains a negative link the putative conclusion must be negative. We cannot make a statement about all members of class A so the conclusion must be particular. Hence a particular negative A-C conclusion is derived. Notice that the direction of the conclusion is a consequence of the direction of the arrows in the combined representation. This, in turn, is dependent on the order of terms in the premises. Hence, the figural effect is neatly explained because the relative positioning of terms is maintained throughout. The flow from end terms, via the b term, is disrupted for figure 2 and 3 syllogisms. Figure 3.2.2 shows, as an example, how the arrows run in opposite directions. Hence, no directionality bias appears in conclusion production.
So far we have seen how possible conclusions are discovered, but we do not yet know whether a putative conclusion necessarily follows. To do this the initial heuristic representation must be modified by attempting to break existing connections by adding alternatives, which are consistent with the premises. If an exhaustive rebuilding programme were undertaken a valid conclusion would eventually be established, or it would be evident that no valid conclusion could hold. However, errors occur because exhaustive attempts to falsify are not made. Consequently error rates on invalid problems are high, and so too are those for certain determinate premise pairs which necessitate exhaustive testing to ensure correct solutions. This pattern of errors is borne out in the data of Johnson-Laird & Steedman and is also suggested in alternative paradigms (e.g. Ceraso & Provitera, 1971; Fisher, 1981). Further support for the analogical model comes from the work of Mazzocco et al. which suggests that people prefer a linear ordering of terms in premises.

Since the analogical model is based on the syllogistic construction task, it represents reasoning in arguably the most unfettered of paradigms. It therefore posits a forward, premise to conclusion, process. We cannot guarantee that such a process would occur on the more restricted versions of the syllogistic
task, however. Despite the success of the analogical model, it has been replaced by an alternative, mental models, explanation. However, the roots of the latter, more general, theory are clearly evident in its predecessor.

Mental models and syllogistic reasoning

As we saw in Chapter 1, the fundamental idea of the mental models theory is that people understand the world by constructing mental models of it in their minds. Figure 3.2.3 illustrates that, unlike the Euler diagram representations of premises proposed by earlier models, the mental models format has the advantage of requiring only one model or diagram to represent the logical properties of each premise type.
Figure 3.2.3 The representation of premises using mental models.

<table>
<thead>
<tr>
<th>Premise</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (x) are (y)</td>
<td>(x = y)</td>
</tr>
<tr>
<td></td>
<td>(x = y)</td>
</tr>
<tr>
<td></td>
<td>(\langle y \rangle)</td>
</tr>
<tr>
<td></td>
<td>(\langle y \rangle)</td>
</tr>
<tr>
<td>Some (x) are (y)</td>
<td>(x = y)</td>
</tr>
<tr>
<td></td>
<td>(x = y)</td>
</tr>
<tr>
<td></td>
<td>((x) \ (y))</td>
</tr>
<tr>
<td>No (x) are (y)</td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>(\langle y \rangle)</td>
</tr>
<tr>
<td></td>
<td>(\langle y \rangle)</td>
</tr>
<tr>
<td>Some (x) are not (y)</td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td></td>
<td>((x) = y)</td>
</tr>
<tr>
<td></td>
<td>(y)</td>
</tr>
</tbody>
</table>

When we draw a conclusion from two premises the number of integrated models we need to construct will vary according to each individual syllogism, but there are never more than three integrated models of the premises to be constructed. As the number of integrated models to be constructed grows, so does
problem difficulty (see Johnson-Laird, 1983, pages 104 and 114). This is seen to be the result of increasing the load on a limited working memory system. As well as storage limitations, working memory is also seen to have processing limitations, so the number of premise manipulations required will also drain resources and hence reduce accuracy levels. In particular, syllogisms in some figures require more adjustments than others. The two principal factors which underlie accuracy are thus (i) number of models which must be considered and (ii) number of premise manipulations required to form models which are easily integrated.

**Premise integration**

Models of premises are most easily integrated by substituting end terms for middle terms which are co-existent. Thus, in the following example, some c's may be substituted for b's; this does not alter the logical meaning of the relationships expressed, but simplifies the integration operation considerably.
Figure 3.2.4 The integration of mental models by substitution of co-existent end and middle terms.

Premise 1 : All A are B
Premise 2 : All B are C

Taking care to construct the same number of linking elements the premises are represented thus:-

Premise 1          Premise 2
a = b               b = c
a = b               b = c
(b)                 (b) = c
                     (c)

Which can be directly integrated by substituting c for b elements thus:-

a = c
a = c
(c)

Some models cannot be integrated by straightforward substitution. While the A-B/B-C figure allows direct substitution of c for b elements, the other three structural relationships of end and middle terms must be rearranged in order to allow
substitution. Johnson-Laird & Bara (1984) report response times for premise pairs which require only one integrated model. They indicate that the A-B/B-C pair takes least time. Then comes the B-A/C-B figure. The B-A/B-C and A-B/C-B figures yield by far the longest response times. Indeed, this is the trend we would expect if we take into account the number and type of operations required. The increase in response time is associated with a decrease in percentage correct over the four figures.

In explaining the well documented directionality bias in drawing conclusions, Johnson-Laird appeals to a 'first in, first out' principle of working memory. If the first end term to be held in model form is 'a' and the second is 'c' (or vice versa), then this is the order in which they should be out-put. In other words, they should appear in that order in the conclusion produced. Notice that, unlike the analogical theory which puts directionality preferences down to a structural feature of the representation used, the mental models theory attributes this to operational characteristics of working memory. This provides internal consistency within the theory.

So far we have explained error rates and latency differences for each figure (at least with respect to recent findings in the construction paradigm); we consider now how error patterns within figure are explained.

The way in which two premises may be integrated is constrained by the mood of each particular premise pair and the difficulty of any given problem is a product of both its mood and figure. Although the theory fares well in predicting the level of accuracy for individual syllogisms, this is based on the number
of models required to solve the problem correctly; it does not explain why specific response patterns occur. Why are some erroneous conclusions associated with specific premise pairs? Take the following example, which requires the construction of three integrated models to form a correct conclusion, and is notoriously difficult.

<table>
<thead>
<tr>
<th>Premise pair</th>
<th>Possible integrated models</th>
</tr>
</thead>
<tbody>
<tr>
<td>All B are A</td>
<td>(1) a (2) a (3) a</td>
</tr>
<tr>
<td>No C are B</td>
<td>a a a</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td>c c = a c = a</td>
</tr>
<tr>
<td></td>
<td>c c c = a</td>
</tr>
</tbody>
</table>

Most subjects seem to construct only model (1); they claim, erroneously, that 'No C are A'. The mental models theory predicts a C-A conclusion; it also suggests that only one integrated model is likely to be formed - but why should model (1) be formed in particular? Subjects rarely, if ever, claim that Some C are A or even All C are A, as they might if models were integrated at random. One possible reason why models may be constructed in a specific order is that the principal aim is to keep within the bounds of conversational maxims (e.g. Grice, 1975). Hence, the first model to be constructed would be the most obvious. This explanation is entirely consistent with the other uses to which mental models are put, e.g. discourse comprehension.
Erickson's (1974; 1978) complete and random combination models. As its title implies, Erickson's complete combination model assumes that each Euler diagram constructed to represent one premise is exhaustively combined with each Euler diagram constructed to represent the other premise. The random combination model assumes that only one, randomly chosen, diagram of each premise is combined with the other to form a composite representation. At the response stage, both models claim that subjects find a label which describes the composite representation(s) derived; however, more than one label may apply to most Euler diagrams (see Appendix I for examples). In order to explain why certain labels are given preference over others, Erickson has to appeal to the atmosphere principle. Johnson-Laird (1983) has pointed out flaws in both the simple and the full-scale models. The former cannot account for the frequently observed endorsement of non-propositional conclusions. The random model always predicts a single composite conclusion - from which a propositional conclusion will always follow. The complete combination model cannot explain why when there is a valid propositional conclusion subjects sometimes claim that there is not. If, on the basis of incomplete encoding, an exhaustive combination occurs on a valid problem there is no way that a propositional conclusion could not be applicable.

We have already noted that neither the atmosphere nor the conversion models can explain the figure effect; unfortunately neither can the set theoretic models. Once the premises are transformed into Euler diagrams, the ordering of terms within the
premises is lost and one figure becomes indistinguishable from any other. (This is also a problem encountered by Mayhew (1981), who has attempted to reformulate Erickson's set theoretic model.) An alternative explanation, which retains the notion of set diagrams but alters their representational form, is the transitive chain theory. It is disappointing that this ingenious replacement also fails to account for figure.

The transitive chain theory.

Sternberg's transitive chain theory (Guyote & Sternberg, 1981; Sternberg & Turner, 1981) is probably the most complex explanation of syllogistic reasoning to date. It is presented in two forms. For our purposes we focus on the performance model. According to this theory, the source of error lies in a failure to combine all premise interpretations, and a biased labelling of combined representations at the conclusion production stage. The proposed representational format is closely related to Euler diagrams. However, because it contains an arbitrary number of set exemplars instead of closed sets, the format is more flexible. However, this type of format produces the same capacity drainage as the Euler circle format, as up to four set relation representations may apply to one premise. When these are combined fully to form the required composite representations the amount of information to be dealt with becomes unwieldy. This then constitutes the first source of error.

The precise form of representation consists of pairs of symbolic "informational components". These are the links from which transitive chains are manufactured.
Figure 3.2.5 The representation of set relations in Sternberg's transitive chain theory.

<table>
<thead>
<tr>
<th>Set Relation</th>
<th>Symbolic Representation</th>
</tr>
</thead>
</table>
| Equivalence  | \( a_1 \rightarrow B \mid b_1 \rightarrow A \)  
               | \( a_2 \rightarrow B \mid b_2 \rightarrow A \)  |
| Subset-set   | \( a_1 \rightarrow B \mid b_1 \rightarrow A \)  
               | \( a_2 \rightarrow B \mid b_2 \rightarrow -A \)  |
| Set-subset   | \( a_1 \rightarrow B \mid b_1 \rightarrow A \)  
               | \( a_2 \rightarrow -B \mid b_2 \rightarrow A \)  |
| Overlap      | \( a_1 \rightarrow B \mid b_1 \rightarrow A \)  
               | \( a_2 \rightarrow -B \mid b_2 \rightarrow -A \)  |
| Disjoint     | \( a_1 \rightarrow -B \mid b_1 \rightarrow -A \)  
               | \( a_2 \rightarrow -B \mid b_2 \rightarrow -A \)  |

By taking one link e.g. \( a_1 \rightarrow B \) \( a_2 \rightarrow -B \) from one pair and combining it with another link, via the middle term, from the other premise e.g. \( b_1 \rightarrow C \) \( b_2 \rightarrow C \) we form one possible composite representation thus,

\[
\begin{align*}
  a_1 & \rightarrow B \\
  a_2 & \rightarrow -B \\
  b_1 & \rightarrow C \\
  b_2 & \rightarrow C
\end{align*}
\]

The formation of A-C or C-A chains follows two set rules which have to be applied repeatedly. Once this is achieved
(apparently faultlessly), all combined representations must be gathered together and redundant or conflicting combinations eliminated. The final stage consists of finding a verbal label to describe the remaining representations. This process is biased by atmosphere and a preference for simplicity whereby preference is given to labels referring to the fewest possible set relations. Notice that we are not told how a subject chooses between A-C and C-A conclusions and why such preferences exist.

Guyote & Sternberg (1981) performed a series of five experiments, designed to test the performance model of the transitive-chain theory. The model was tested with abstract, thematic, categorical and conditional reasoning tasks. Analyses indicated that the model accounted for a considerable proportion of variance in the data. However, systematic, unexplained variance was also observed - a significant proportion of which was attributed to the figural effect.

A large proportion of variance was accounted for by atmosphere assumptions. The adoption of existing biases like atmosphere should not necessarily be criticised; other theorists also do this. However, we should consider to what extent a new theory relies on the incorporation of such biases. If we take away the atmosphere component the transitive chain model is weakened considerably.

So far we have considered the transitive chain model with regard to abstract problems. Guyote & Sternberg have made the additional claim that the model can account for performance with various types of problem content. Testing the transitive chain model, they found consistent differences between problem
contents; abstract (letters of the alphabet), and concrete (factual, counter-factual and anomalous). Factual problems were associated with significantly more correct responses than the other types of content. Moreover, this difference was claimed to be due to the greater probability of combining more than one pair of premise representations. Although the model predicts that the more combinations the greater likelihood of a correct response, it does not explain why factual content should promote an increase in combinations.

The transitive chain model is a complex and unparsimonious explanation, which takes as its principal components the relatively successful aspects of established theories (i.e. Erickson's failure to combine all premise interpretations and Woodworth & Sells' atmosphere bias). The novel elements introduced fail to improve on the descriptive adequacy of predecessors, and add a certain implausibility to the model, which makes it untenable in a practical sense.

3.3 Error due to incorrect translation of derived conclusions.

As we have seen, a large proportion of theories totally neglect the figure effect. Apart from atmosphere, which is a non-logical explanation, other theories (which assume some form of deductive mechanism) use figure to their own ends. Such theories, like those based on illicit conversion, suggest that the figure of a syllogism is distorted during encoding. Other theories obviously overlook figure because of the method of representation used (e.g. Erickson's set theoretic model and Sternberg's transitive chain theory). Any theory which posits logical inference must account for the role which figure plays,
hence it must also explain the figure effect. We now consider one such theory put forward by Dickstein (1978a) in which people are assumed to encode and combine premises logically.

Dickstein's (1978a) theory of syllogistic reasoning is based on an earlier theory proposed by Frase (1968). In order to explain the former we need to run through the latter. Frase noticed that the arrangement of terms in the four figures corresponded to four distinct learning paradigms (see Appendix I page 5). On this basis he suggested that syllogistic reasoning may be influenced by mediated associations analogous to those observed in verbal learning research. By chaining terms together via the middle or linking term (M), the S-M-P chain formed from figure 1 corresponds with the order of end terms appearing in the conclusion. Thus, by a process of forward chaining one arrives at the appropriate conclusion. For figure 4 syllogisms the premises form a P-M-S connection which must be reversed in order to correspond with the order of terms in the given conclusion. This requires backward chaining which is relatively harder than its counterpart, forward chaining. Hence, more errors are predicted in figure 4 than figure 1 syllogisms. Since no chain can be directly formed from the premises of figure 2 and 3 syllogisms, performance is neither aided nor hindered by the relative positioning of terms.

Frase's theory is unacceptable for a number of reasons. Most obvious is the failure to explain the nature of the mechanism. Furthermore, it is descriptively inadequate because it completely neglects response variations due to mood. In this respect, Dickstein's (1978a) explanation is an improvement.
Dickstein's directional processing theory.

According to Dickstein, subjects always draw the logically correct conclusion and the direction of that conclusion depends on the positioning of terms in the premises. If the end terms appear in the same half of the conclusion as they do in the premises, then any propositional conclusion derived will always be valid. This is termed forward processing and applies to figure 1 syllogisms. If, as for figure 4 syllogisms, the positioning of end terms in premises and conclusion conflict, the subject processes the premise information in a backwards (P-S) fashion. This is where error creeps in, because subjects erroneously assume that by converting the conclusion derived they will make it comparable with the (S-P) conclusion(s) given.

Dickstein makes two specific predictions. Firstly, as Frase predicted, figure 1 syllogisms will be easiest, figure 4 the hardest and figures 2 and 3 intermediate (because either forward or backward processing may be induced here). The second prediction qualifies the first and distinguishes directional processing from mediated association predictions. It states that the errors observed will depend upon the mood of conclusion derived with backward processing. If the conclusion is logically convertible, then no errors should occur on figure 4 syllogisms. If the converse of the conclusion derived is not equivalent to the valid conclusion given then specific errors will be found.

In a test of this theory, Dickstein carefully controlled for premise conversion, using EI premise pairs, which yield a valid (O) conclusion in all figures, because both premises are
convertible. With backward processing such premises yield no valid conclusion, however. Hence, the theory predicts the endorsement of the 0 and nonpropositional conclusion on figures 1 and 4 respectively. The complementary set of IE premise pairs was also included. This premise pair yields no valid conclusion with forward processing, but leads to an 0 conclusion with backward processing. Equally revealing were those premise pairs for which forward and backward processing yield the same conclusion. For these problems no figural effect (defined in terms of error) should occur. In general, Dickstein's findings supported the theory's predictions. The effect of figure was not significant for those problems on which forward and backward processing lead to identical conclusions. However, as Evans (1982a, page 101) points out, the model is described as an algorithm yet its predictions are confirmed only as statistical tendencies. On figure 2 an erroneous response was given to one fifth of problems with a valid (S-P) conclusion. However, according to Dickstein, no errors should occur here. In addition, on figure 4 problems, as few as one third of the responses given were erroneous. The theory is further weakened by the presence of unpredicted error patterns on figure 4 syllogisms.

The directional processing theory is a relatively parsimonious attempt to explain a complex phenomenon. Unfortunately, the theory is both descriptively and explanatorily inadequate. It can only describe relative error rates and fails to explain exactly how subjects transform figures 2 and 3 in order to process them in a directional (S-P or P-S) manner. Do subjects convert the premises in order to do this? If they do so
we would expect this to be an additional source of error. Premise conversion leads us to the question of conclusion conversion. There is no evidence to suggest that subjects selectively convert the conclusion. Indeed, there is no direct evidence to show that subjects consistently convert any proposition when part of a syllogism.

Theories which consider errors at each stage of the reasoning process.

In the preceding sections, the theories considered focused on one, sometimes two, stages in reasoning as the major source of error. We now consider two explanations: Newell's (1981) problem space hypothesis and Fisher's (1981) three-factor model, which allow error to creep in at various stages in reasoning.

Newell's (1981) application of the problem space hypothesis.

Newell's account of syllogistic reasoning focuses on ideal performance and does not attempt to explain errors directly. The inferential procedure is based on a Venn diagram analysis, where each component is represented by a symbolic label. For example:

<table>
<thead>
<tr>
<th>Syllogism</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All A are B</td>
<td>Nec A+B+, Poss A-B+, Poss A-B-</td>
</tr>
<tr>
<td>No C are B</td>
<td>Nec C+B-, Poss C-B+, Poss C-B-</td>
</tr>
<tr>
<td>Therefore: Some A are not C</td>
<td>Nec A+C-</td>
</tr>
</tbody>
</table>

With the syllogistic evaluation task, means ends analysis is used as the basis of integration in an attempt to form an instance of the goal state (i.e. the conclusion). Since a
possible instance does not ensure necessity, there is a search for contradictory combinations. From the absence of contradiction, necessity is inferred.

This procedure works well enough for the evaluation task and one could easily adapt it to account for error - say at the premise labelling or combination stage or in the search for contradiction. However, it is not clear how subjects would behave in the syllogistic construction or multiple choice tasks. The weaknesses of the problem space approach were discussed in chapter 1. For these reasons we must conclude that the problem space hypothesis, as it stands, can make only a limited contribution to our theoretical understanding of syllogistic reasoning.

Fisher's (1981) three factor model

Fisher has attempted to isolate specific stages of the reasoning process in order to establish their relative contribution to error. In a series of experiments subjects were required to (i) interpret individual premises with respect to a selection of other propositions, (ii) select from a list of alternatives the proposition which described the combination of modified premises and (iii) decide whether each of four given conclusions was permissible according to a variety of response alternatives. On the basis of (i) above, existing models of premise misinterpretation did not fare well. No single theory could account for the large degree of individual differences in interpretation. (Only one subject interpreted each premise in a strictly logical manner). Using modified premises it appeared that a large number of subjects had difficulty in deciding which
premise combinations could apply. Interestingly, responses varied according to the number of response alternatives available. Several explanations could be offered for this (see chapter 2).

From the data collected in each separate experiment, Fisher concludes that error is attributable to three stages in syllogistic reasoning: premise interpretation, combination and conclusion selection. From the experiments conducted by Fisher there is indeed evidence that error exists at such stages - when each stage is considered in isolation. By dissecting the traditional syllogistic task in this way, Fisher runs the risk of changing its nature because each operation is taken out of context. Evidence exists to suggest that individual premises, for instance, may be interpreted differently when they appear as one of a pair (Begg & Harris (1982), Experiment 3). To this problem we must add another complication. Immediately after completing the premise interpretation task, Fisher's subjects were given the multiple choice syllogistic reasoning task. Under such conditions we cannot guarantee that the interpretation of individual premises did not influence their subsequent interpretation when they appeared as one of a pair of premises. Since Fisher does not present his data in sufficient detail we cannot tell whether the syllogistic task yielded anything like traditional response profiles.

It is difficult to assess the usefulness of Fisher's findings because of the problems outlined. Although Fisher is clearly aware of the influence which paradigm may have on performance, he seems to overlook the possibility of distortion due to dissecting and then presenting the syllogistic task in
Conclusions and implications for the study of belief bias.

The majority of theories of syllogistic reasoning are both explanatorily and descriptively inadequate. Theories of premise misinterpretation each have their own peculiar shortcomings. Almost all of them cannot explain figural effects. This is not to say that premise misinterpretation does not occur. Indeed, as we have seen, several studies suggest that it does, but no single theory can account for the interpretations observed. Theories which place the principal source of error at the combination stage provide by far the most acceptable accounts. Empirical evidence suggests that, in a syllogism by syllogism analysis, the most difficult tend to be those which require exhaustive combination of premise encodings in order to arrive at the logically correct conclusion. Several theories point to the incomplete combination of premises. However, in aspects of detail some of these theories are clearly more adequate than others. The transitive chain model is highly complex and fails to account for figural bias - so too do Erickson's set theoretic models.

Clearly, theories must not only account for error, but also for correct reasoning. Models which propose logical premise interpretation and combination, such as the problem space hypothesis simply cannot cope with the level of systematic errors observed. Dickstein's directional processing theory faces the same problem, even when illogical conclusion conversion is taken into account.

The introduction to this chapter stated that a theory of
syllogistic reasoning must explain both mood and figure effects, as well as the relationship between difficulty and degree of necessary premise combination. The analogical model can account for all these variables, but lacks the generalisability of its successor, the mental models theory. Overall, the latter is the most adequate and plausible explanation, but it still leaves some questions unanswered.

As far as belief bias effects are concerned, the majority of theories of syllogistic reasoning steer clear of this issue. Theories which rely on logical premise encoding and inference rules to explain correct responses have difficulty in explaining content based effects. Newell's problem space hypothesis allows for alterations which could include some form of heuristic basis for belief bias. However, problems of testability limit the usefulness of the problem space approach. Dickstein's directional processing theory would have to attribute belief bias to error at the conclusion selection stage. However, illicit conclusion conversion alone could not account for belief bias. Theories based on premise misinterpretation offer a prime opportunity for rationalists to argue against the suggestion that reasoning per se is belief biased. Revlin (1978) has extended his conversion plus feature selection theory claiming that belief bias is the result of selective conversion blocking by realistic material and we consider this explanation in the next chapter. There are no theories of belief bias which attribute performance to incomplete premise combination, although there seem to be promising possibilities. It could be argued that salient, i.e. real-life, relationships are established before any other combination. We
might expect unbelievable (and hence less salient) relationships to be considered last of all and hence, if we are dealing with a limited capacity system such relationships may not be considered at all. This would result in the endorsement and production of believable as opposed to unbelievable conclusions. In contrast to this, Guyote & Sternberg have actually claimed that factual content results in an increased likelihood of combining more than one pair of premise representations. Since they do not provide sufficient detail of the specific content used and how this was combined with problem structure it is impossible to accurately assess how the two may have interacted. Before we can evaluate the suitability of existing models and decide upon the type of modifications required we need to look in detail at studies of belief bias in order to document the exact nature of the effect. The next chapter is devoted to this end.

Note 1
Revlis has since changed his name to Revlin
# Chapter 4

The belief bias effect: A characterisation and search for determining factors.

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>The syllogistic evaluation task.</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Do beliefs influence reasoning?</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>What is the nature of belief bias and what are its determinants?</td>
<td>80</td>
</tr>
<tr>
<td>4.2</td>
<td>The multiple choice/response selection task.</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Do beliefs influence reasoning?</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>What is the nature of belief bias and what are its determinants?</td>
<td>92</td>
</tr>
</tbody>
</table>

Belief bias: general conclusions. | 98   |
In reviewing individual studies of belief bias it is useful to distinguish between the paradigms used, bearing in mind the psychological differences which may be associated with each. In addition studies may be categorised according to the questions they set out to answer.

Most studies have sought to establish whether or not a relationship exists between a belief (stated or assumed) and the evaluation of a given, controversial, conclusion. One study, reported by Revlin & Leirer (1978) and Revlin, Leirer, Yopp & Yopp (1980), has investigated the effect of belief in the premises, using neutral conclusions. In this case no correspondance was found between belief and response. The remaining studies have not controlled for premise believability. Although no evidence exists to suggest any such effect, it cannot be ruled out on the basis of one study.

Two types of syllogistic paradigm have mainly been used; the evaluation and response selection (or multiple choice) task. Notably, little attention has been paid to the syllogistic construction task. This remains the case despite appeals for such studies (e.g. Madrzycki, 1978) and the widely held assumption, in social psychology, that subjects should show more attitude change under such conditions (Jones & Gerard, 1967).

The following review examines studies by categorising them according to the paradigm used and their investigative aim.
4.1 The syllogistic evaluation task

Do beliefs influence reasoning?

One of the earliest experiments designed to answer this question was carried out by Lefford (1946). This study was run during the war years and made use of current issues of the time. Lefford's conclusions concerned the behaviour of Nazis, the value of war and politics. Such topics would presumably have aroused strong beliefs during this period (see also Morgan & Morton, 1944).

In Lefford's experiment, subjects were asked to evaluate the validity of two sets of syllogisms. One set contained 'controversial' content; the other set contained what was presumed to be neutral content. After completing this task subjects were asked to go back through the problems answered and indicate, on the same sheet, whether or not the conclusions given were believable. Subjects were run in four groups, which differed according to the order in which controversial and neutral problems were given. Unfortunately, the relationship between validity and believability could not be controlled because the latter was not established before administration of the test.

Results were presented according to a graphical analysis. Accuracy scores for neutral syllogisms, presented first, were distributed normally with the greatest part of the population falling around the mean of the range. The distribution of accuracy scores for controversial syllogisms followed a somewhat different, J shaped distribution. There was a pronounced piling up of scores at the lower end of the scale. Bearing in mind that
the problems used in the neutral and controversial sets were identical in structure and length, the differing distributions provide strong support for the contention that belief influences judgement. Clearly, this has a detrimental effect on reasoning with controversial conclusions. Furthermore, differences between distributions according to whether neutral or controversial problems were presented first or second suggested carry-over effects. When controversial syllogisms were presented first, reasoning on subsequent neutral syllogisms was detrimentally affected. The opposite effect seemed to occur when controversial followed neutral problems, in this case a slight improvement was seen on the latter problem type. When neutral and controversial problems were randomly intermingled no such effects were observed. On the basis of these results, Lefford proposes that some form of set effect operates when problems of differing content are presented in blocks, which cannot become established under conditions of random presentation.

Although Lefford's findings thus far seem clear cut, there is one additional finding which could question the further interpretation of results. In trying to establish the relationship between belief and reasoning response, Lefford observed not only a relationship between belief and response on controversial items, indicating that believable conclusions were accepted and unbelievable rejected; he also found an identical relationship on 'neutral' items. This effect is almost undoubtedly due to the forced choice of truth or falsity imposed for neutral items. Subjects could not answer that they had no belief either way - as one would expect with neutral syllogisms.
Forced to make a somewhat meaningless decision, subjects may have based their decision of truth status largely on the answer which they gave to the reasoning task. This would create an apparent concurrence between belief and reasoning response, and explain why reasoning accuracy was not apparently influenced by the belief expressed in this case.

Unfortunately, the method used by Lefford to assess beliefs may also have indirectly influenced the responses given to controversial syllogisms. This could have occurred in either of two ways. Firstly, subjects may have been influenced by the letters 'T' and 'U', representing 'True' and 'Untrue', which were printed by the side of each problem. This may have encouraged subjects to use belief as a criterion, when otherwise subjects may not have entertained such a consideration. Secondly, each subject had ample time to alter his reasoning response so that it was consistent with his subsequent belief rating. This is perhaps less likely than the first explanation, but is certainly not inconsistent with the predictions made by theories of attitude change.

A recent attempt to replicate Lefford's findings (Nehrke, 1972), found a significant difference between responses to controversial and neutral problems, but the J-shaped distribution found by Lefford was not replicated. Instead, responses to both neutral and controversial problems followed normal distributions, but the mode was higher in the case of neutral problems. No carry over effects were observed. Nehrke did not apparently administer an attitude questionnaire, but assumed that about half the valid and half the invalid problems were
believable, the remainder being unbelievable. If this is the case, and belief acts as Lefford suggests, then we would expect to find normally distributed accuracy scores. In the case of Lefford's experiment the equal distribution of believable and unbelievable conclusions over valid and invalid syllogisms was not obtained. This may well account for the observed difference in the distribution of accuracy scores between the two studies.

An additional component of Nehrke's study supplied information concerning individual differences in reasoning with controversial content. In particular he reports that highly educated subjects fare better than the less well educated on both controversial and neutral syllogisms, which substantiates the findings of Feather (1964; 1967).

An effect of belief and a possible interaction with educational level was reported by Thouless (1959). This study contained syllogisms concerning socialism, life-after-death and war. Using a rating method similar to that used by Lefford, subjects were asked to work through the test paper, marking the conclusions given by circling T(true), F(false) or (?) according to their believability. This, in contrast to Lefford's study, was performed before the syllogistic task, on the same sheets of paper. Thouless gives no details of the structure of problems used, but reports that a 'simple' syllogistic test given to mature students yielded approximately 40% error, of which 73% of responses were in the direction of belief. A more 'difficult' test given to university students yielded no overall effect of belief despite the fact that the materials included were chosen to excite strong beliefs. The university students had an error
rate of only 10% on the simple version and 16% on the harder version. It is concluded that either the higher educational level of university students led to less reliance on belief (also suggested by Nehrke, 1972) or that the university students guessed the aim of the experiment and reacted by rejecting a priori knowledge. Certainly, comments made by this group suggested that they had guessed the aim of the experiment; however, it may be unwarranted to claim that such subjects could consciously and deliberately resist the influence of belief. In interpreting this study it should be noted that Thouless gave examples of sound and unsound arguments which both had definitionally true conclusions. He also explained the response categories, 'sound' and 'unsound'. It is possible, therefore, that the university students grasped the task requirements more so than the mature students.

What is the nature of belief bias and what are its determinants?

One of the most carefully designed studies to investigate the nature of belief bias was that of Janis & Frick (1943). They hypothesized that the belief bias effect would be manifest as follows:

1) If there is agreement with the conclusion, more errors will be made in judging logical validity by accepting invalid arguments than rejecting valid arguments.

2) If there is disagreement with the conclusion, more errors will be made in judging logical validity by rejecting valid arguments than by accepting invalid arguments.

Four types of problem were used:
1) valid problems with believable conclusions  
2) valid problems with unbelievable conclusions  
3) invalid problems with believable conclusions  
4) invalid problems with unbelievable conclusions.

Attitudes were ascertained by post-experimental questionnaire. Although attitude ratings and reasoning responses do not appear to have been written on the same sheet, it is nevertheless possible that responses to the first test may have influenced responses to the second test. This seems unlikely, however, as data are presented as if an even, predictable split of believable and unbelievable ratings were observed (see Table 4.1.1).

Table 4.1.1. The distribution of errors on each of the four problem types used by Janis & Frick (1943) (taken from table 2 page 76).

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Believable</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Unbelievable</td>
<td>22</td>
<td>11</td>
</tr>
</tbody>
</table>

Chi square analyses indicated a significant difference between categories. Further tests on believable and unbelievable syllogisms independently indicated a significant difference in both cases. It was, therefore, concluded that the results obtained indicated two tendencies; the influence of belief and disbelief. Clearly, such a conclusion is unwarranted because no neutral syllogisms were included for comparison. We can say, however, that when logic and belief agree, fewest errors are observed.
A similar experiment by Kaufman & Goldstein (1967) expanded upon Janis & Frick's findings and reported a logic x belief interaction. Kaufman & Goldstein investigated a number of factors; belief, validity and quantification. In a well designed study, the believability of conclusions was ascertained by asking a group of subjects (who did not take part in the experiment proper) to indicate their belief in the statements. This method is preferable to that of asking the same subject to indicate belief and solve problems concerning the same statements. Statements with which over 80% of subjects agreed were used as PA (positive affect) statements, NA (negative affect) statements were collected similarly. An equal number was taken from each category and a number of neutral (unrated) statements, comprising one third of their combined total, was also added. A within subject design was employed. Results indicated a strong effect of validity along with several interaction effects. Such effects are difficult to interpret. For existential statements, overall, the number of errors on neutral syllogisms was significantly greater than those on believable (PA) and unbelievable (NA) syllogisms. Predictions concerning order effects were only upheld on conclusions with existential, as opposed to universal conclusions. The order effect was such that the erroneous rejection of valid arguments was N>NA>PA, and the erroneous acceptance of invalid arguments was PA>N>NA.

The interaction between logic and belief can be interpreted in two ways. There is either a more marked belief bias effect on invalid than valid syllogisms, or there is a greater effect of logic on unbelievable than on believable conclusions. A triple
interaction between logic, belief and type of quantification suggests that an error of acceptance is more likely on existential than universal conclusions when the conclusion is believable. As Kaufman & Goldstein point out, the fact that they found strong validity and triple interaction effects shows that no sweeping statements can be made about errors, affective loading or quantification by themselves. This observation emphasizes the necessity of controlling for such factors, and the importance of reporting experimental design in detail. Both are unfortunately neglected by many studies.

Thus far studies have suggested that belief bias is manifest as a tendency to accept and reject conclusions according to their believability. Since we are rarely wholeheartedly in agreement or disagreement with a given statement we should consider the effect of varying degrees of belief. Feather (1964; 1967) set out to investigate this question.

Feather's experiments focused on three particular individual differences: strength of belief, reasoning ability and intolerance of inconsistency. The controversial syllogisms he used concerned the Christian religion and were randomly intermingled with neutral syllogisms. All problems were matched according to logical form and differed according to content only. A religious attitude scale was administered one week after the syllogistic task. Form E of the Dogmatism Scale, developed by Rokeach (1960); a test for 'open' and 'closed' minds, and a test of intolerance of ambiguity developed by Budner (1962), which measures ambiguity as a source of perceived threat and unambiguity as desirable, were also administered.
On the basis of ratings on the religious attitude scale, subjects were divided into two groups, one pro-religious and one anti-religious. The difference in numbers between groups was, however, quite large; 131 versus 34 respectively. For the pro-religious group, correlations indicated that evaluation of syllogisms in a manner consistent with attitude was positively related to intensity of attitude, negatively related to level of critical ability and positively related to intolerance of inconsistency (assuming that the measure used was an index of this variable). A negative correlation was observed between intolerance of ambiguity and critical ability. Similar relationships were predicted, but not generally found, for the anti-religious group. A tendency was observed for subjects with relatively high negative religious attitude scores to make more errors in the anti-religious direction. This tendency was not significant. Relatively few subjects were categorisable as strongly anti-religious, however. Consequently this may have reduced the strength of the predicted correlations. This explanation seems likely in the light of a later study (Feather, 1967). In this study subjects were preselected. A small number (10) of pro-religious society members were selected from each of three societies, an equal number of subjects were selected on the basis of their strong atheist beliefs. All but one previous finding was replicated for the pro-religious groups. (The positive relationship between syllogism evaluation in accordance with attitude and intolerance of ambiguity fell just short of significance.) For the anti-religious group all but one of the former predictions were upheld. Once again the predicted
relationship between syllogism evaluation and religious attitude fell just short of significance although the trend was clearly in the right direction.

Taken together, the results of both studies indicate that the stronger the belief, in whatever direction, the more likely it is to influence reasoning on controversial syllogisms - although this is more marked for pro-religious belief. The higher the level of critical ability, the less likely it is that belief will intrude. The more intolerant of ambiguity a person is, the more likely he is to exhibit belief-based reasoning.

So far the effect and nature of belief bias seem consistent across studies. However, a series of experiments conducted by Frase (1966a,b; 1968) appear to challenge the characterisation developed thus far. At the outset, we should note that Frase's own findings are perplexingly inconsistent. Frase (1966a) found that belief was correlated with incongruity; such that the more incongruous an assertion, the less credible it was. On this basis Frase suggested that, in terms of a mediation model, belief in an assertion is contingent upon the incompatibilities of the words associated and even more critically upon the numbers of relations which are associated within a given hierarchy (as stated by quantifiers). In an initial test Frase asked subjects to complete an unspecified test of syllogistic reasoning. The statements used were either universal or particular and of either low, moderate or high incompatibility (see Frase, 1966b, for a detailed account of how incompatibility was measured). One week after attempting the syllogistic test, subjects were asked to rate the conclusions according to their believability. Frase reports that the effect
of highly incongruous conclusions on error scores was less than the effect of moderate incongruity. A later study appears to produce slightly different findings (Frase, 1966b). In this case there was a significant main effect of affective terms and the largest number of errors was found with moderate incongruity. This differed significantly from error scores for low incompatibility. Error scores for high incompatibility problems did not differ significantly from either low or moderate. In addition there was a significant difference between error scores on universal and particular conclusions, where more errors occurred on the latter. However, no incongruity x quantifier interaction was observed. Frase (1968) reports somewhat different results. Although performance with quantifiers followed the same pattern (despite training on their interpretation), the U shaped relationship between incompatibility and reasoning accuracy was absent. Instead, both low and medium levels differed significantly from high, which had the highest level of percentage errors. Since Frase gives no detailed description of the syllogisms used - save that only 25% were valid - this would suggest that incompatibility led to the acceptance of invalid arguments. Such an effect is directly contrary to that reported by previous studies. However, this aspect of Frase's findings does not appear to be consistent within his own range of experiments. Indeed, the only consistent finding is that greater errors are observed on particular as opposed to universal conclusions. We should, therefore, await replication before placing too much reliance on Frase's other findings.

The studies considered so far have all used a syllogistic
evaluation paradigm. With the exception of Frase (1966a,b; 1968) all imply that responses are predominantly in accordance with belief in such cases. However, the interaction between logic and belief indicates that the belief bias effect is not isolated. Either it is stronger on believable problems or stronger on invalid problems. Furthermore, belief based evaluation is also linked to strength of belief and intolerance of ambiguity. It also seems that belief bias is negatively related to critical ability. All in all, belief bias does not seem to be the result of guessing the correct conclusion. Such an explanation is too simplistic. In order to achieve a better understanding of the effect we need to look at a greater variety of studies. We now go on to consider those which employ the response selection paradigm.

4.2 The multiple choice/response selection task.

Do beliefs influence reasoning?

A relatively weak effect of belief bias was reported by Wilkins (1928). In the main experiment, subjects were given a mixture of single premises and premise pairs, followed by three conclusions. Some problems were valid; others contained logical fallacies. There were four types of problem content:

A) Familiar: problems with thematic content, whose conclusion had no apparent everyday truth status.

B) Symbolic: problems containing letters of the alphabet.

C) Unfamiliar: problems containing very unfamiliar scientific terms or nonsense words which were selected to resemble scientific terms.

D) Belief bias: in this case problems were accompanied by
conclusions whose assumed truth status conflicted with their validity. Hence, valid and invalid conclusions were unbelievable and believable respectively.

Although Wilkins reports no statistical analysis, Evans (1982a page 106) reports his own analysis of the data which shows performance on thematic problems to be highly significantly superior to that on symbolic and unfamiliar problems, but only marginally significantly superior to that on belief bias problems. Comparison in terms of percentage correct show that familiar content improves performance, but that this improvement is reduced when beliefs conflict with logical validity.

The relatively small difference between categories A (thematic) and D (Belief bias) can be attributed to at least two causes. Firstly, Wilkins did not obtain independent ratings of believability. All conclusions were classified by Wilkins - not the subjects themselves. The believability judgements of subjects could have been at odds with those of the experimenter in some cases. Secondly, categories A and D were not absolutely discrete in terms of content. In some cases category A conclusions had real world truth status, for example: All children who are not good will be neglected by Santa Klaus and All people who have intellectual curiosity make use of reference libraries. The failure to ascertain believability ratings and the overlap of contents may well have acted to reduce the difference between accuracy scores on A and D categories.

In a similar study, Morgan & Morton (1944) gave subjects abstract syllogisms (containing letters of the alphabet) and structurally identical syllogisms with conclusions of a
controversial nature. Examining each individual pair of problems, paired according to logical structure, Morgan & Morton compare response distributions over the five conclusions given for the abstract and controversial problems. Noting differential distributions, the authors conclude that this is due to selection on the basis of atmosphere in the former case and inherent believability in the latter. The effect of belief was characterised as a tendency to choose those conclusions with which one agreed.

Several aspects of Morgan & Morton's study may be severely criticised. Firstly, no independent measure of belief was used. The authors classify responses in a *post hoc* fashion according to their own assumptions concerning the beliefs, wishes and fears of subjects. On these criteria it is highly likely that a 'belief' biased explanation could be offered for almost any response.

A second problem is that often more than one conclusion to any given problem was believable. Sometimes one conclusion is preferred, at other times popular choice is less distinct. The authors offer no explanation as to why certain believable conclusions are preferred in some cases and, in others, why no single, believable conclusion is chosen more than any other.

Another problem with this study is the dubious categorisation of selections on the basis of atmosphere. In syllogism 11, for example, the combination of particular negative and universal negative premises is claimed to produce a universal negative atmosphere; clearly this is not the case. Equally suspect are the atmosphere predictions (peculiar to Morgan & Morton) for premises containing quantifiers such as 'most' and
'usually'. In such cases it is not altogether clear exactly what atmosphere would predict, if anything.

Overall, Morgan & Morton appear to suggest that belief overrides the otherwise strong effect of atmosphere. Close examination of the tables given, showing the distribution of selections for each problem when expressed in both abstract and controversial terms, indicates no systematic alterations of choice. Although there are significant alterations, they do not always reflect a shift from one response to one particular type of alternative. Alteration of responses undoubtedly occurs; whether belief is the determining factor may be debatable.

The studies of Gordon (1953) and Henle & Michael (1956) produce strong counter claims to those of Morgan & Morton. Both studies asked subjects to state which conclusions were believable or unbelievable. Gordon concludes that atmosphere is more dominant than belief in conclusions. Henle & Michael actually attempted to replicate Morgan & Morton's study. They presented structurally identical problems, but altered their content to take account of temporal changes in topic of controversy. Instead of conclusions concerning people of India, for example, they included some referring to Communists - a rising American concern, yet the conclusion favoured was identical in form to that reported by Morgan & Morton. It seems, therefore, that conclusion choice was independent of its content. Two further experiments in the same series also fail to attribute response profiles to belief. Certain features of all three experiments may have contributed to this finding. Firstly, in their replication of Morgan & Morton's study, Henle & Michael did not establish the
degree of belief in conclusions. Subjects were simply asked to state whether or not the various conclusions were true in their opinion. Given Feather's (1964; 1967) findings, the lack of belief bias could be due to an absence of strongly held attitudes. In their second experiment, which notably contained structurally simpler problems and a preponderance of valid syllogisms, subjects were asked to rate their attitude towards Communism and Russia in general. Since all thematic problems were concerned exclusively with this topic, subjects may well have guessed the purpose of this experiment and deliberately avoided selection of attitudinally consistent conclusions. In their third experiment subjects were trained in the logic of syllogisms and allowed to use diagrams; this would probably make structure a highly salient feature of the task.

In a more recent study Madrzycki (1978) also failed to establish any (conventional) belief bias effect. Using syllogisms with pro or anti religious statements as their conclusions, Madrzycki found only 6.8% to 10% of conclusions on all problems to correspond to beliefs. As in Gordon's (1953) study, a marked atmosphere effect was observed. An unpredicted finding was that significantly more errors were made when the valid conclusion was believable. This means that subjects were avoiding the believable conclusion and not only selecting a relatively unbelievable conclusion, but also an invalid one because, apparently, no indeterminate syllogisms were used. Since the data are not reported in detail it is not possible to discover whether this tendency was specific or widespread. The fact that responses can be distinguished on the basis of the believability of the logical
conclusion suggests that there is an effect of belief. However the effect is not characterisable as a tendency to accept and reject those conclusions with which one agrees and disagrees respectively. It is noteworthy that this effect occurred even though all subjects had, at some time, been trained in logic.

What is the nature of belief bias and what are its determinants?

One of the most notable opponents to the suggestion that a non-logical effect, based on belief, either replaces or distorts the logical reasoning process is Revlin (Revlin & Leirer, 1978; Revlin, Leirer, Yopp & Yopp, 1980). Revlin has extended his conversion model of syllogistic reasoning (see chapter 3) in order to provide an alternative, logic-based interpretation of what appears to be an irrational tendency. From the standpoint of his model, Revlin argues that deductive errors on categorical syllogisms are only indirectly affected by the truth value of conclusions and do not reflect insufficiencies in the reasoner's logical skills.

The extension of the conversion model is best explained with reference to example syllogisms. The following example demonstrates how a logically incorrect response, which is consistent with belief, can be explained by the original conversion model.
All Russians are Bolsheviks.

Some Bolsheviks are undemocratic people.

Therefore:

(a) All undemocratic people are Russian;
(b) No undemocratic people are Russian;
(c) Some undemocratic people are Russian;
(d) Some undemocratic people are not Russian;
(e) None of the above is proven.

Only one of the two premises above is logically convertible; the universal affirmative (A) premise is not convertible, the particular affirmative (I) premise is convertible. According to Revlis's (1975) conversion model it is postulated that the 'A' statement will be interpreted as set equivalence. In other words, subjects will understand "All Russians are Bolsheviks" to mean also that "All Bolsheviks are Russians". Logically, this assumption cannot be made. If this assumption is made, as Revlin claims, the logically prescribed conclusion is altered, from conclusion (e) to conclusion (c).

To extend the assumptions of the conversion model to categorical reasoning with other types of material, Revlin et al. attempted to show that decisions can be predicted from independently assessed premise encoding. The model claims that when the premises contain information about categories that are already available to the reasoner, long-term memory may provide working memory with more information than may have been contained or intended in the presented material. It follows from this that, although with abstract content all premises (and sometimes
conclusions) are converted, this may not necessarily be the case if long-term memory excludes certain conversion interpretations. This process is termed 'conversion blocking' and acts to prevent the representation of semantically deviant relations.

In one experiment subjects were categorised according to the answers given to a questionnaire concerning a) the believability of statements and b) their convertability. Statements appearing in the questionnaire were then incorporated into problems as syllogistic premises. The premises to such problems ranged from the innocuous to the controversial; however, conclusions were always neutral. Revlin categorised subjects as believers or non-believers, and then sub-divided them into converters and non-converters. Results indicated that responses corresponded closely to premise interpretation rather than premise believability. Although this implies that interpretation and not belief in the premises is an influential factor when reasoning with controversial materials, crucially it does not rule out the proposition that belief in the conclusion is a powerful determinant of responses. The second experiment was designed to refute this claim.

In the second experiment, problems consisted entirely of syllogisms whose premises were legally convertible. In each case illegal premise conversion could not differentially affect logical decisions. A range of believable and unbelievable conclusions were presented, but according to the conversion model, belief in the conclusion should not be related to selection rates. In other words, reasoning in this experiment should be entirely logical. Incorrect decisions should,
therefore, be ascribed to irrational or non-logical processes rather than premise misinterpretation.

In this experiment twice as many determinate (EI-1, EI-2) syllogisms as indeterminate (II-1, II-2) syllogisms were used. For half of the determinate premise pairs the logical conclusion was categorised as believable, the remaining half, unbelievable. Since categorisation was not based on independent believability ratings, the classifications made were of dubious status. The following example problem, from Revlin et al's experiment, demonstrates this point. The logically prescribed conclusion (marked by an asterisk) is deemed by the authors to be unbelievable, although this conclusion is probably not the most believable of the selection presented, it is arguably not unbelievable.

No U.S. governors are members of the Harem Club.

Some Arabian sheiks are members of the Harem Club.

Therefore:

(a) All Arabian sheiks are U.S. governors.
(b) No Arabian sheiks are U.S. governors.
(c) Some Arabian sheiks are U.S. governors.
(d) Some Arabian sheiks are not U.S. governors.*
(e) None of the above is proven.

In all cases atmosphere favoured the logical conclusion to determinate premise pairs, regardless of its believability. Conversely, for indeterminate premise pairs, the conclusion categorised as believable was never favoured by atmosphere. Such
believable conclusions were always universal (either affirmative or negative) and premise pairs produced a particular affirmative atmosphere. On such problems belief was opposed by both logic and atmosphere.

Results indicated an effect of belief, such that responses were less accurate when logic and belief conflicted. An additional manipulation to investigate differences between empirical and definitional truth status showed no significant differences in response patterns. There was, however, a significant interaction which indicated that accuracy deteriorated more in those cases where logic conflicted with a definitionally true conclusion. In those cases where logic conflicted with definitional truth, the dominant error (62.5% of all errors) was due to the choice of the non-propositional conclusion. Clearly, this cannot be seen as a conventional effect of belief bias. This behaviour could indicate that, in the face of conflict between logic and belief, neither belief nor logic wins and the response elicited represents a decision to avoid conflict by selecting the non-propositional conclusion.

It is clear that this experiment shows a somewhat unconventional effect of belief. Revlin argues that the effect is limited, in that although the effect was significant at the 1% level, reasoning accuracy was otherwise impressively high. Taking into account the fact that in most cases atmosphere worked against belief, which in turn was categorised somewhat dubiously, the high level of accuracy is only superficially impressive. Since twice as many determinate as indeterminate problems were used, and logically prescribed propositional conclusions were
always favoured by atmosphere, reasoning accuracy on such problems would be expected to be relatively high.

Although this experiment is presented as evidence for the influence of premise misinterpretation, it probably constitutes one of the strongest pieces of evidence in support of belief bias on the response selection task. In the light of the criticisms made, the conversion model does not provide an adequate explanation of results. The fact that the belief bias effect may not have been conventional does not detract from its importance; on the contrary it may constitute something of a revelation.

On balance, evidence from the response selection/multiple choice task is in favour of some form of belief based effect, although it seems less clear-cut that that from the evaluation task. The one study which argues for belief bias, Morgan & Morton (1944), is clearly flawed, but then so too is the opposing study of Henle & Michael (1956). Although Madrzycki (1978) and Revlin et al (1980) argue against any effect of belief, their results suggest otherwise. Both studies show unconventional effects of belief. It is difficult to interpret Madrzycki's findings, but Revlin's results suggest that conflict between logic and belief leads to a 'conflict avoidance' response, whereby a neutral, non-propositonal conclusion is preferred.
Belief bias: General conclusions

In evaluating the evidence for and against belief bias we are disadvantaged by having to restrict our observations to the evaluation and response selection/multiple choice paradigms. It is evident from chapter 3 that valuable insight may be gained from studies of the syllogistic construction task which are, unfortunately, lacking in this particular area of research. From the restricted range of experiments available we arrive at the following conclusions.

1. Belief bias is evident on both types of syllogistic task, although its exact nature is not clear.

2. Most evaluation task studies suggest that belief bias may be manifest as a tendency to accept and reject conclusions according to whether they are believable or not. Although Frase (1966a,b, 1968) presents evidence which might imply that this relationship is not so straightforward. Experiments which employ the response selection/multiple choice task seem to be less supportive of the belief bias effect. Although some studies claim that no belief bias exists, others show uncharacteristic effects which do not conform to those generally found on the syllogistic evaluation task.

There is clearly no definitive study of belief bias to rely on, but it is true to say that the evaluation task has provided the most satisfactory and compelling evidence of belief bias to date. In order to develop our understanding, these studies, in particular, should be replicated and extended. Since we know relatively little about belief bias and how it operates we cannot attempt to apply theories of syllogistic reasoning to its...
explanation. Suffice to say that premise misinterpretation alone does not seem to hold the key. Detailed investigation must first be carried out to ascertain the nature of belief bias and the relative strengths of logic and belief.
SECTION TWO

EXPERIMENTS

CHAPTER FIVE
Belief bias and the syllogistic evaluation task. 101

CHAPTER SIX
Belief bias and the syllogistic evaluation task. 147

CHAPTER SEVEN
The logic x belief interaction - the search for an explanation. 193
Chapter 5

Belief bias and the syllogistic evaluation task.

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Introduction</td>
<td>102</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>104</td>
</tr>
<tr>
<td>Method &amp; Design</td>
<td>105</td>
</tr>
<tr>
<td>Protocol scoring procedure</td>
<td>109</td>
</tr>
<tr>
<td>Results &amp; Discussion</td>
<td>110</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>115</td>
</tr>
<tr>
<td>Method &amp; Design</td>
<td>116</td>
</tr>
<tr>
<td>Results &amp; Discussion</td>
<td>119</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>126</td>
</tr>
<tr>
<td>Method &amp; Design</td>
<td>127</td>
</tr>
<tr>
<td>Results &amp; Discussion</td>
<td>129</td>
</tr>
<tr>
<td>Conclusions &amp; General Discussion</td>
<td>136</td>
</tr>
</tbody>
</table>
INTRODUCTION

The three experiments which comprise this chapter focus on belief bias and the syllogistic evaluation task. Studies in this area of research have mainly used the syllogistic evaluation or multiple choice task. One reason for choosing the former type of task is that belief bias appears to be more consistent across studies which employ this mode of presentation. However, there are several problems with such studies and it is necessary that their findings be substantiated before any attempt is made to expand upon them. A second reason for using this type of task is that several of the methodological problems incumbent upon the multiple choice task can be avoided.

In order to add greater depth to this initial series of experiments, verbal protocols were recorded and analysed in an attempt to relate them to the reasoning response given. Both Henle (1962) and Scribner (1975) have analysed verbal protocols. They seem to have assumed that protocols reveal the underlying cause of behaviour and have argued, largely on this basis, that belief bias is due to task rejection or misconception and distortion of problem premises. Accordingly, their results appear to be supportive of rationalist claims. However, the interpretation of protocols is a matter of debate. Using a different reasoning task, Wason & Evans (1975) claimed that such reports did not reveal insight into the origins of responses apparently based on non-logical bias. In terms of the dual process theory, which they produced, verbal processes are seen as part of a qualitatively different reasoning process from that
which forms the basis of decisions. Moreover, Wason & Evans claimed that responses attributed to non-logical bias are rationalised by the subject. Hence, protocols reveal only the purported cause of behaviour, attributed (retrospectively) by the subject in an attempt to bring some order to his behaviour. Verbalisations are thus seen to be divorced from non-logical thought processes; of which belief bias is possibly an instance. Evans (1984) has since altered his theoretical interpretation of reasoning, but still rejects the notion that protocols reveal privileged access. Along similar lines, Nisbett & Wilson (1977) have argued that protocols frequently reflect expectations or theories which subjects have concerning the cause of their behaviour. However, alternative views also exist. For example, Ericsson & Simon (1980) have argued that verbal reports may reveal the focus of attention and constitute causal accounts, but this depends on the type of task and the method of collection used. Within the bounds of Experiments 1 to 3, it is difficult to assess which characterisation is most appropriate. However, whilst bearing in mind these conflicting views, the collection of verbal protocols may still constitute a valuable source of additional data.
EXPERIMENT 1

The aim of the first experiment was to test the relative weighting given to logic and belief in syllogistic reasoning. Revlin et al (1978,1980), using the multiple choice task, have claimed that the belief bias effect is relatively weak compared with that of logic. Furthermore, bias is attributed to premise misinterpretation and not any non-logical process. Several criticisms of Revlin et al's experiment were made in Chapter 4. For the most part, flaws were unavoidable, since they were an inevitable consequence of the response selection paradigm employed. However, this may have led to an overestimation of people's logical abilities. To overcome this problem Experiment 1 employed an alternative paradigm. The syllogistic evaluation task was chosen to provide greater control for the potentially confounding effects of atmosphere and distortion due to competing conclusions of similar belief status. Both are problems which may be encountered with other paradigms. In order to avoid error due to premise conversion (as proposed by Revlin's model) all premise pairs were convertible. Hence, if any belief bias was present, it could not be attributed to this particular type of premise misinterpretation.

Experiment 1 began the investigation of protocols by examining retrospective verbalisations and relating them to responses given to problems which systematically varied according to the believability and logical status of the conclusion.
METHOD

Materials

Half of the syllogisms presented were valid and half were invalid. Problems were in the form of BIO-2 syllogisms, for example

No highly trained dogs are vicious,
Some police dogs are vicious.
Therefore: Some highly trained dogs are not police dogs.

(Invalid, Believable conclusion)

No police dogs are vicious,
Some highly trained dogs are vicious.
Therefore: Some police dogs are not highly trained.

(Invalid, Unbelievable conclusion)

(Full details of problems used are given in Appendix II)

Half of the valid conclusions presented were believable and half were unbelievable and this was also true for invalid conclusions. There were thus four types of problem.

Since problems consisted of E and I premise pairs, both
premises were 'legally' convertible, which means that the terms within such statements may be reversed without altering their meaning in logic. Since the major and minor premises were always of this type, even if subjects did convert the premises as Revlin et al suggest, this could not in itself produce logical errors in this experiment.

B and I premises produce an atmosphere which favours an O conclusion. In the present experiment all conclusions given for evaluation were of this type. Hence, all conclusions, both valid and invalid, were favoured by atmosphere bias. This could not, therefore, confound any comparisons between problems.

To reduce the artificiality of the task, problems were embedded in prose passages, which were approximately 80 words in length. Four different types of passage content were used. Each took the form of a current affairs article. The four topics were: a) public response to the behaviour of police dogs; b) the provision of aid for third world countries; c) attempts to reduce the number of people smoking cigarettes and d) the relationship between wealth and hard work (see Appendix III for examples of each passage). The following is an example of passage type (a).

'Dogs are used extensively for the purpose of guarding property, guiding the blind and so on. No highly trained dogs are vicious, however, many people believe that their temperament cannot be trusted. The police service use dogs a great deal in their work. Some police dogs are vicious and although fatal accidents are rare, there is still growing concern over their widespread use.'
'If the above passage is true, does it follow that:

SOME HIGHLY TRAINED DOGS ARE NOT POLICE DOGS?'

The above conclusion is invalid, but believable. The conclusion is deemed invalid because the stated relationship between highly trained dogs and police dogs does not necessarily follow from the two problem premises embedded in the prose passage. The pragmatic implications of the passage itself should have no bearing on the evaluation of the conclusion given, since they are logically irrelevant. In the same way, a priori knowledge about the truth status of the conclusion is logically irrelevant. Subjects are instructed to accept the statements within the passage to be true. They are asked to evaluate the logical status of the conclusion only. According to standard propositional logic, the answer to the above problem is that the given conclusion is invalid. Therefore, the logical response is to reject the conclusion given.

The believability of all valid (0) conclusions was ascertained by questionnaire. The questionnaire design was based on the Likert (1932) procedure. Conclusions were rated on a seven-point scale, ranging from 1: completely unbelievable to 7: completely believable. A rating of 4 thus represented the judgement that a statement was 'neither believable nor unbelievable'. Subjects who took part in the questionnaire study were excluded from the experiment proper. This method is preferable to that of obtaining questionnaire ratings and problem evaluations from the same subjects since this might affect
reasoning performance or vice versa. The believability ratings for conclusions used in this experiment are given in Appendix IV.

Design

Each subject received each of the four types of prose passage, solving all four problem types in all (valid problems with believable and unbelievable conclusions, invalid problems with believable and unbelievable conclusions). Combination of problem type and passage content was such that each content occurred with each problem type an equal number of times. Presentation order was randomised.

Subjects

Twenty-four undergraduates at Plymouth Polytechnic acted as paid volunteers. They had no previous experience of this task and were tested individually.

Apparatus

Tape recorder (visible to subject).

Procedure

Task and Instructions

The instructions and problems for each subject were presented on typed cards. All problems were presented individually and each problem card remained in front of the subject for reference when explaining decisions. The instructions were as follows:
This is an experiment to test people's reasoning ability. You will be given four problems. In each case you will be given a prose passage to read and asked if a certain conclusion may be logically deduced from it. You should answer this question on the assumption that all the information given in the passage is, in fact, true. If you judge that the conclusion necessarily follows from the statements in the passage, you should answer "yes", otherwise "no".

Please take your time and be sure that you have the right answer before stating it. When you have decided, I will then ask you to explain why you believe the conclusion to be valid or invalid as the case may be. Any questions?

Subjects' protocols were recorded on a tape recorder for later analysis.

Protocol scoring procedure

Each protocol was scored according to two criteria:

(i) Presence or absence of a reference to both of the logically relevant premises.
(ii) Presence or absence of references to irrelevant information from within the passage (ie anything other than the two premises) or from extraneous sources (eg real-life information).

Both types of reference were classed as irrelevant because the conclusion should be evaluated according to its relationship to the premises making statements about the two conclusion terms. Logically, the relationship between these terms may only be deduced by combining the two terms via a common (middle) term; which appears twice only - once in each of the two problem premises.

RESULTS AND DISCUSSION

The percentage frequencies of subjects accepting the conclusion, ie deeming the argument to be valid, are shown for each type of problem in Table 5.1. It is clear from this table that a marked belief bias is present, ie a tendency, over all problems, to accept more believable than unbelievable conclusions ($p < 0.001$, one-tailed sign test). More valid than invalid conclusions were accepted overall, ($p < 0.02$, one-tailed sign test) and a significant belief x validity interaction was present ($p < 0.05$, two-tailed sign test) The nature of this interaction is consistent with the findings of Kaufman & Goldstein (1967). There is, in fact, a significant effect of belief on valid problems also ($p < 0.01$, one-tailed sign test).
These results conflict with the predictions of the conversion model. The claim that previously reported belief bias effects may be due to an artifact of uncontrolled premise conversion is not applicable to this experiment, in which all problem premises were legally convertible. Moreover, the results of this experiment refute the suggestion made by Revlin et al that belief bias is relatively weak compared with rational processes. They report 83% correct when logic accorded with belief and over 67% correct when logic conflicted with belief. The corresponding percentages in the present experiment are 92% and 27%. This substantiates the claim that their methodology led to an inaccurate picture of the relative effects of belief bias and logical processes.

Subjects were classified into those responding to logic, belief or the characteristic combination of logic and belief shown by the interaction. Only one subject's responses accorded entirely with logic, whereas eleven subjects responded entirely according to belief. Seven subjects responded according to logic on valid problems and belief on invalid. (The five unclassifiable subjects were two subjects who accepted all conclusions and three who had uncategorisable overall response patterns). Some individual differences clearly exist for this task; interestingly very little pure logical ability is evident and those subjects who are influenced by belief fall into two categories - those who respond to belief only, and those who also seem to be 'sensitive' to the logic of the problem.

Turning to protocol analyses, Table 5.2 presents the classification frequencies of protocols scored on the two
Table 5.1

Percentage frequency of subjects accepting the conclusion given. Divided according to the four problem types. Experiment 1 (n=24)

<table>
<thead>
<tr>
<th>Believable</th>
<th>Unbelievable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>92</td>
</tr>
<tr>
<td>Invalid</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 5.2

Classification frequencies for protocols on the two criteria, broken down by problem type and response given. Experiment 1 (n=24)

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Response</th>
<th>Both premises</th>
<th>Irrelevant Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>NR</td>
</tr>
<tr>
<td>Valid-</td>
<td>Yes</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Believable</td>
<td>No</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Valid-</td>
<td>Yes</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Unbelievable</td>
<td>No</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Invalid-</td>
<td>Yes</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Believable</td>
<td>No</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Invalid-</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Unbelievable</td>
<td>No</td>
<td>2</td>
<td>20</td>
</tr>
</tbody>
</table>
criteria; reference to both premises and reference to irrelevant information.

Protocol content can be viewed in two ways. It can a) reflect the information forming the basis of the response or, b) reflect a justification or rationalisation by the subject, made in the light of a response given. Both of these characterisations lead to the same predictions concerning the relationship between a response (based on logic or belief) and protocol content (logically relevant or irrelevant information). On problems where logical and belief-based reasoning lead to the same response we cannot tell on what basis the response has been made and hence cannot falsify the verbalisation produced. The most interesting problem types are those for which logic and belief conflict. Here, it is easier to discern the basis of a decision and relate this to the protocol content. Both protocol characterisations predict that responses which are consistent with logic (but not belief) will be associated with more references to logically pertinent information. Responses which are consistent with belief (but not logic) will be associated with more references to factual or at least logically irrelevant information.

The predicted response x protocol interactions were tested using Fisher Exact probability tests. For each of the four problem types, two tests were carried out, one for premise scores and one for irrelevant information scores. The four cells on which each test was based were, for example; presence or absence of reference to premises x 'Yes' or 'No' response. This yielded $2 \times 2 = 4$ cells on which the test was performed. Only one of the eight Fisher Exact tests yielded a significant interaction. (This
was due to the lack of a sufficiently even split between 'Yes' and 'No' response frequencies for the majority of remaining problem types. The predicted interaction between response and reference to irrelevant information was upheld and significant for premises followed by a valid-undebatable conclusion ($p < 0.05$).

An interesting pattern shown in Table 5.2 indicates that only slightly more than half of the references to premises are associated with the correct response. This may present problems for the rationalist interpretation of protocols, since premises are restated by subjects, but nevertheless more than one third of these subjects produce the logically incorrect response. Given that, in such cases, premises are restated without logical distortion (which may lead to incorrect responses) it is difficult to explain the frequency of associated incorrect responses unless a logically distorted reasoning process is postulated. However, it could be argued that such errors arose because of misinterpretations which may not be apparent in protocols.

The protocol data collected have provided a disappointing lack of information concerning the relationship between protocols and response. Since this relationship is further investigated in the following experiments, discussion of the present findings will be deferred until the end of this chapter.
EXPERIMENT 2

This experiment was designed to replicate and extend the findings of the previous experiment. It aims to clarify several aspects of the interpretation given to Experiment 1. A major difference between Experiment 1 and most other studies of belief bias is that problem premises were embedded in a prose passage. It could be argued that this type of presentation format could increase belief bias by diverting attention from the logically critical premises. Hence, Experiment 2 included a group of subjects who received unembedded premises. Another difference is the instruction to verbalise, which could affect responses. Consequently, Experiment 2 incorporated a group of subjects who were not requested to verbalise. To further investigate the relationship between response and object of reference in protocols another group was added who were asked to 'think aloud' whilst attempting to solve each problem. Ericsson & Simon (1980) have argued that this type of protocol is more likely to yield an accurate picture of the focus of attention and hence the information heeded, than is the retrospective report requested in Experiment 1.

Finally, the problem structures were modified to take account of 'figural bias' as discussed by Johnson-Laird & Steedman (1978). In Experiment 1, validity was confounded with the order of terms within conclusions. Valid conclusions ran from C to A, and invalid ran from A to C (where A and C are the first and last terms to appear in the premises, other than the linking term (B)). According to Johnson-Laird & Steedman's findings, though not their model, there is a slight bias in preference for
C-A conclusions with the premise types employed in Experiment 1. This may have artificially augmented the level of logical performance observed. Experiment 2 includes both valid and invalid problems which are equally associated with A-C and C-A conclusions.

METHOD

Materials

The syllogisms and prose passages were the same as those used in Experiment 1, except for the structural changes made to create an equal number of A-C and C-A conclusions for both valid and invalid problems. This was achieved by interchanging the quantifiers of each of the original premises and reversing the conclusion. For example:

Original syllogism (valid) Control syllogism (valid)
No A are B Some A are B
Some C are B No C are B
Therefore: Some C are not A Therefore: Some A are not C

Design

As in Experiment 1, all subjects received four problems consisting of all four problem types combined with all four passage contents. In this experiment four subject groups were used.

The first group received premises embedded in prose passages
and was required to explain, retrospectively, the reason for their decision, as in Experiment 1. This group is represented mnemonically as 'Rverb(P)', where Rverb stands for 'retrospective verbalisation', and (P) indicates premises embedded in a prose passage. The remaining groups are labelled according to the same principle.

The second group was required to verbalise in the same manner as the former group, but received problem premises (followed by a conclusion) which were not embedded in any prose passage. This group is simply labelled 'Rverb'.

The third group, labelled 'Cverb(P)', received premises embedded in prose passages and was required to verbalise concurrently, i.e. to 'think aloud' whilst attempting to solve the problem.

The fourth and final group, 'Nverb(P)', received premises embedded in prose passages, but unlike other groups, subjects were not required to verbalise at any stage.

Each of these groups was then subdivided into two further groups, one of which received only A-C conclusions for both valid and invalid problems, and the other received C-A conclusions only.

Subjects

64 undergraduates at Plymouth Polytechnic acted as paid volunteers. They had no previous experience of this task, or any training in logic. They were tested individually.

Apparatus
Tape recorder (visible to subject).

Procedure

The instructions and problems were presented in the same manner as in Experiment 1.

The instructions were as follows:

Group

Rverb(P) - Instructions as for Experiment 1.

Rverb - Instructions as for Experiment 1, except that any reference to the prose passage was omitted.

Nverb(P) - Instructions as for Experiment 1, except that the request to verbalise was omitted.

Cverb(P) - Instructions as for Experiment 1, first paragraph only, they continued:

'Whilst you are trying to solve each problem I would like you to try to 'think aloud' as much as you can. Please do not let this distract you from the task in hand - which is to obtain the correct solution to the problem. If, at any time during the task, I do not think that you are speaking enough I will simply prompt you to speak a little more.

Please take your time and be sure that you have the right answer before stating it. Any questions?'

As in Experiment 1, subjects' protocols were recorded on a tape recorder for later analysis. Protocols were scored using the same procedure as in Experiment 1.
RESULTS AND DISCUSSION

The percentage frequency of subjects judging arguments to be valid is shown in Table 5.3.

Clearly the order of terms in the conclusion had no discernible effect on responses, therefore, further analyses were collapsed over this factor. Table 5.4 presents the percentage frequency of subjects accepting arguments, broken down by group.

All groups exhibit trends which replicate those observed in Experiment 1. In order to identify differences between groups, a set of four 2 x 4 Chi-square tests were carried out comparing yes/no frequencies across all four groups on each problem type. None of these analyses yielded a significant result.

Binomial tests on the combined data of all groups (n=64) yielded highly significant effects of a) belief (p < 0.001); more believable than unbelievable conclusions were accepted, b) logic (p < 0.001); more valid than invalid conclusions were accepted and c) an interaction between logic and belief (p < 0.01); where belief bias was more marked for invalid than valid syllogisms (All one-tailed tests). A strong effect of belief was, however, observed on valid problems (p < 0.001). Overall, subjects were correct 87% of the time when logic accords with belief and 48% of the time when it conflicted with belief.

There are aspects of the present findings that are crucial to the interpretation of Experiment 1. Firstly, since no differences were observed between performance on embedded and unembedded problem premises, this indicates that the belief bias observed in Experiment 1 was not due to the irrelevant material
Table 5.3

Percentage frequency of subjects accepting arguments as valid, broken down by problem type and order of terms in conclusion. Experiment 2 (n=64)

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bel</td>
<td>Unbel</td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-C</td>
<td>88</td>
<td>56</td>
</tr>
<tr>
<td>C-A</td>
<td>84</td>
<td>68</td>
</tr>
<tr>
<td>x</td>
<td>86</td>
<td>62</td>
</tr>
</tbody>
</table>

Bel = Believable  Unbel = Unbelievable

Table 5.4

Percentage frequency of subjects in each group accepting arguments as valid, Experiment 2 (n=16 in each group)

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bel</td>
<td>Unbel</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rverb(P)</td>
<td>81</td>
<td>63</td>
</tr>
<tr>
<td>Rverb</td>
<td>81</td>
<td>69</td>
</tr>
<tr>
<td>Cverb(P)</td>
<td>87</td>
<td>50</td>
</tr>
<tr>
<td>Nverb(P)</td>
<td>94</td>
<td>67</td>
</tr>
</tbody>
</table>

Note: Bel = Believable, Unbel = Unbelievable
distracting subjects from the logical task. Secondly, since no significant differences exist between groups which verbalised and that which did not, the requirement to verbalise seems not to have distorted responses in Experiment 1. Finally the lack of any response preference for C-A as opposed to A-C conclusions, indicates that the confounding of conclusion directionality and validity did not lead to an overestimate of the logical component in this task.

The results of protocol analyses for each group are shown in Table 5.5. Inspection of this table suggests that the distribution of classification frequencies is very similar for all three groups. This similarity was assessed by rank ordering of the 16 cell frequencies for each group and applying Kendall's coefficient of concordance (cf Siegel, 1956). There was high and significant concordance for both mention of the premises ($W = 0.792, p < 0.001$) and mention of irrelevant information ($W = 0.871, p < 0.001$). Consequently further analyses were performed on the combined data for all three groups.

The analyses revealed highly significant interactions between the answer given to the problem and the protocol classifications for the Valid-Unbelievable problem. Subjects accepting the valid conclusion against its believability made more references to the logical premises ($X^2 = 24.61, p < 0.001$) and fewer references to irrelevant information ($X^2 = 15.11, p < 0.001$).

An interaction was also observed on the Invalid-Unbelievable problem. Only six subjects went against both logic and belief to accept the argument as valid. However, all
Table 5.5

Classification frequencies for protocols of Experiment 2 (n=16 in each group) broken down by response given.

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Both Premises</th>
<th>Irrelevant Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RvP</td>
<td>Rvb</td>
</tr>
<tr>
<td>Val-Bel</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>R</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>NR</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Val-Unb</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>R</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>NR</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Inv-Bel</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>NR</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Inv-Unb</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>NR</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

Note: -
R = reference; NR = no reference;
Y = yes response; N = no response;
RvP = Rverb(P); Rvb = Rverb;
CvP = Cverb(P); Comb = Combined
six referred to both logical premises and none referred to irrelevant information. Fisher exact probability tests revealed a significant interaction with the majority 'no' responders in each case ($p < 0.005$ and $p < 0.05$, respectively). The simplest interpretation of these findings is that these subjects ignored beliefs and reasoned from the premises, but they did so with faulty logic.

The two consistent interactions found for the Valid-Unbelievable problem clearly uphold the predictions made by the dual process model and Ericsson & Simon. The accompanying prediction, concerning the interaction between response and protocol on the other conflict problem (Invalid-Believable) is not borne out, however. Notice that, on this problem type, the premises do not actually contradict the Invalid-Believable conclusion. Perhaps subjects who respond according to belief have no qualms about citing the premises here because they are not contradicting themselves. They do not cite the premises when rejecting a Valid-Unbelievable conclusion though, because then they would be introducing contradiction. This would explain why we find the predicted interaction for the Valid-Unbelievable problem, but not the Invalid-Believable problem. If this state of affairs did apply, this would mean that some subjects, responding according to belief, actually understood the logic of the problem.

So far protocol analyses have concentrated on the content of protocols; whether or not the premises are mentioned and whether or not irrelevant information is included. This may indicate what subjects attend to, particularly in the case of concurrent
verbalisation, but it does not indicate how the problem is approached. An alternative method of analysis is to categorise protocols according to the order in which problem components are mentioned. Hence, in the case of concurrent verbalisation, the flow of attention can be identified. This method of analysis is a central technique in the examination of 'thinking aloud' protocols (see Newell & Simon, 1972) and may be useful in identifying heuristics, i.e. factors which guide the reasoning process. Although, in the present experiment, only 16 subjects were requested to 'think aloud', initial investigative analysis suggests that the majority of those subjects primarily focus attention on the conclusion and a relatively small number of these subjects then go on to mention one or both problem premises. Consequently it seems that less than half of the subjects in the thinking aloud group actually reason in a 'forward' premise to conclusion, fashion. Clearly, such limited data constitute an inadequate basis for anything but tentative interpretation. In order to examine further the suggested tendencies it is necessary to run another experiment to replicate and extend the findings relating to the concurrent verbalisation group.

Taken together Experiments 1 and 2 have shown consistent results for reasoning profiles and protocols. The pattern of acceptance rates across the four problem types is particularly fascinating. There is a curious interaction here. When logic and belief agree, correct performance is at a high level, but when logic and belief conflict this level drops considerably. Performance on the Valid-Unbelievable conflict problem seems
better than on its counterpart, the Invalid-Believable problem. Why should this be so? Does this pattern reflect a greater sensitivity to logic on valid problems or could it be that unbelievable problems encourage or aid logical analysis? Before elaborating upon these hypotheses we must rule out a possible confounding factor which may have influenced logical performance. Experiment 3 addresses itself to this question whilst taking a closer look at concurrent verbalisations.
EXPERIMENT 3

Controls were employed in Experiment 2, which indicated that a possible confounding of conclusion directionality and validity could not have artificially augmented the level of logical performance. However, it is still possible that the pattern of correct responding is an artifact of an alternative response bias. In all the problems used so far, the quantifier 'some' always modified the same term in the premises as it did in the valid conclusion presented. For example:

\[
\text{No A are B} \\
\text{Some C are B} \\
\text{Therefore: Some C are not A}
\]

For invalid problems this was not the case. For example:

\[
\text{No A are B} \\
\text{Some C are B} \\
\text{Therefore: Some A are not C}
\]

Clearly, some form of feature matching bias could account for the effect of validity here.

This possibility can be investigated by employing figure 3 syllogisms in addition to figure 2 as control problems. The following problem is logically equivalent to the valid syllogism given above, the only difference being its figure.
No B are A

Some B are C

Therefore: Some C are not A

Since, in all figure 3 syllogisms, the two terms (A and C) appearing in the conclusion stand as predicates in the premises, the proposed response bias could not operate (unless the I premise was converted). Experiment 3, therefore, compared performance on figure 2 with that on figure 3 syllogisms.

The second modification employed in this experiment concerned the emphasis placed on logical necessity in the instructions. Henle (1962) has claimed that belief bias may be the result of rejecting the logical task. Although an inspection of protocols had showed no direct support for this contention, logical necessity was given greater emphasis in instructions to reduce such a possibility.

METHOD

Design

Each subject received eight problems, each embedded in a different prose passage. Four problems were figure 2 and four were figure 3 syllogisms. The four problems in each figure consisted of the four problem types used in Experiments 1 and 2. Four further scenarios were constructed to add to the four used in previous experiments. Each subject received each of the eight scenarios, randomly matched to the eight problems (with the constraint that each problem type occurred an equal number of times with each content). Presentation order was randomised. All
subjects were instructed to 'think aloud' when solving problems.

Details of additional prose passages are given in Appendix III. The conclusion believability ratings for additional problem contents are given in Appendix IV.

Subjects

Thirty-two first year Psychology students at Plymouth Polytechnic participated in partial fulfilment of a course credit requirement. All were tested individually. Subjects had no previous experience of this task nor any training in logic.

Apparatus

Tape recorder (visible to subject).

Procedure

Problems were presented in the same manner as described in Experiment 1. The relevant section of the modified instructions was as follows:

'Your task is to decide whether or not a given conclusion follows logically from the information given - and this information only. You must assume that all the statements within the passage are true - this is very important. If, and only if, you judge that the given conclusion logically follows from the statements given in the passage you should answer "yes", otherwise "no".

The final sentence of this extract was repeated at the very end of the instructions.
Protocols were tape recorded and subsequently transcribed and analysed as in previous experiments.

**RESULTS AND DISCUSSION**

The percentage frequencies of subjects accepting conclusions for both figures is given in Table 5.6.

From Table 5.6 it is evident that no differences exist between performance on figure 2 and figure 3 problems. The feature matching explanation of the validity effect was, therefore, rejected. Subsequent analyses were performed on combined data.

Binomial tests revealed that significantly more believable than unbelievable conclusions were accepted \((p < 0.001)\) and significantly more valid than invalid conclusions were accepted \((p < 0.001)\). The interaction, while in the same direction as observed previously, fell just short of significance \((p = 0.067)\) (All one-tailed tests). Subjects were correct 93% of the time when belief agreed with logic and 43% of the time when belief conflicted with logic.

The results of this experiment have, therefore, clarified the interpretation of Experiments 1 and 2. Firstly the feature matching explanation of the validity effect has been rejected on the basis of comparisons between performance on figures 2 and 3. Secondly, additional emphasis given to the concept of logical necessity in instructions produced identical effects of logic and belief, and an interaction which followed the trend, shown in previous experiments. Assuming that the instructions given were effective, this suggests that the observed belief bias effect and
relative levels of logic and belief were not due to a rejection of the logical task, although we cannot rule this out entirely. Furthermore, an examination of the frequency distribution of responses for each problem type for each content indicated that the observed effects are not attributable to any specific content or contents.

We turn now to protocol analyses, which were performed in the same manner as those of Experiments 1 and 2. Table 5.7(a) sets out the data according to protocol classification and response given for each problem type and each figure.

Protocol analyses on the data shown in Table 5.7(a) showed no significant interactions between reference to premises and response, although a trend is observed for the Valid-Unbelievable condition. No significant interactions between reference to irrelevant information and response were observed.

Clearly it is difficult to derive a conclusive interpretation from these results. There are, of course, other ways of looking at the same data. Table 5.7(b) shows the probability of giving a particular explanation as a function of the response made. This is the appropriate way to look at the data if one assumes that they reflect rationalisations.

In addition to the suggested interaction on the Valid-Unbelievable problem, there appears to be a tendency for subjects accepting the Invalid-Believable problem to make more references to irrelevant information. If these data are interpreted as reflecting rationalisations then this implies that subjects perceive a conflict between logic and belief in this condition also.
Table 5.6

Percentage frequency of subjects accepting arguments as valid, divided according to figure and problem type. Experiment 3 (n=32)

| Problem Type | Valid | | Invalid | | |
| | Bel | Unbel | Bel | Unbel |
| Figure 2 | 91 | 53 | 69 | 3 |
| Figure 3 | 91 | 53 | 66 | 9 |
| Combined | 91 | 53 | 67 | 6 |

Note: Bel = Believable; Unbel = Unbelievable

Table 5.7(a)

Classification frequencies for protocols of Experiment 3 (n=32). Broken down by the response given to each problem type.

| Problem Type | Response | Both Premises | | Irrelevant Information | | |
| | | Fig 2 | Fig 3 | R | NR | R | NR | R | NR | R | NR |
| Valid- Believable | Yes | 5 | 12 | 16 | 7 | 17 | 12 |
| | No | 1 | 7 | 1 | 7 | 9 | 6 |
| Valid- Unbelievable | Yes | 24 | 8 | 13 | 8 | 9 | 9 |
| | No | 20 | 8 | 11 | 8 | 9 | 6 |
| Invalid- Believable | Yes | 11 | 10 | 14 | 10 | 14 | 7 |
| | No | 18 | 4 | 11 | 7 | 5 | 6 |
| Invalid- Unbelievable | Yes | 18 | 3 | 12 | 0 | 2 | 1 |
| | No | 22 | 7 | 22 | 2 | 16 | 13 |
Table 5.7(c) considers the data from a different viewpoint. This table gives the probability of responses, given the protocol scores. This is the appropriate way to look at the data if one assumes that protocols reflect the basis on which subjects were reasoning.

The pattern of responding shown in Table 5.7(c) reveals a protocol distribution which is clearly compatible with that predicted by Ericsson & Simon. For the two conflict problems, subjects correctly accepting the Valid-Unbelievable conclusion are more likely to have referred to the premises. On the other hand, subjects who accept the Invalid-Believable conclusion, against logic, are more likely to have made no reference to the premises. This suggests that subjects who produce the logical conclusion are more likely to have focused on the premises and subjects responding according to belief are more likely to have paid little or no attention to the premises.

If protocols are assumed to reflect the basis of responses, a useful method of analysis is to code the order in which problem components are mentioned. This not only provides evidence concerning the focus of attention, but also the 'approach' employed by subjects. Such an analysis may, for instance, differentiate between premise to conclusion ('forward'), conclusion to premise ('backward') or other types of directional processing. The preceding analyses treat each type of protocol as identical. Clearly, the different types of protocol may reflect fundamentally different approaches. Protocols were divided into three categories: premise(s) to conclusion, conclusion to premise(s) and conclusion only. For the first two categories
### Table 5.7(b)

Percentage of references to premises and irrelevant information as a function of the response given to each problem type.

**Experiment 3 (n=32). Both figures combined.**

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Response</th>
<th>Both Premises</th>
<th>Irrelevant Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid-</td>
<td>Yes</td>
<td>27</td>
<td>57</td>
</tr>
<tr>
<td>Believable</td>
<td>No</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td>Valid-</td>
<td>Yes</td>
<td>64</td>
<td>47</td>
</tr>
<tr>
<td>Unbelievable</td>
<td>No</td>
<td>36</td>
<td>53</td>
</tr>
<tr>
<td>Invalid-</td>
<td>Yes</td>
<td>23</td>
<td>61</td>
</tr>
<tr>
<td>Believable</td>
<td>No</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Invalid-</td>
<td>Yes</td>
<td>50</td>
<td>75</td>
</tr>
<tr>
<td>Unbelievable</td>
<td>No</td>
<td>25</td>
<td>72</td>
</tr>
</tbody>
</table>

### Table 5.7(c)

Percentage of 'Yes' responses as a function of protocol scores.

**Experiment 3 (n=32).**

<table>
<thead>
<tr>
<th>Both Premises</th>
<th>Reference</th>
<th>No Reference</th>
<th>Reference</th>
<th>No Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94</td>
<td>89</td>
<td>94</td>
<td>86</td>
</tr>
<tr>
<td>Valid-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believable</td>
<td>67</td>
<td>39</td>
<td>50</td>
<td>56</td>
</tr>
<tr>
<td>Valid-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbelievable</td>
<td>50</td>
<td>75</td>
<td>72</td>
<td>61</td>
</tr>
<tr>
<td>Invalid-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believable</td>
<td>12</td>
<td>4</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Invalid-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbelievable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
protocols can refer to either one or both premises. This method of scoring does not take references to irrelevant information into account. Table 5.8 sets out the protocol data, according to the order in which problem components are mentioned.

In Experiment 2, protocol analyses of this form identified two basic strategies, which can be labelled 'forward' and 'backward' reasoning strategies. The former is characterised as an initial focusing of attention on the logical premise(s). The latter strategy consists of focusing attention on the conclusion only, or on the conclusion and, subsequently, the logical premise(s). Table 5.8 indicates that the majority of protocols can be thus categorised. Furthermore, protocols appear to be diagnostic of the amount of logical reasoning undertaken. Premise to conclusion protocols are associated with the least belief bias in the two conflict conditions. Those focusing on the conclusion first and then the premise(s) show intermediate levels of belief bias. As we might expect, those protocols focusing on the conclusion only show the most belief bias of all. However, even these people appear to be sensitive to logic. Notice the drop in belief based responses on the two conflict problems, compared with no conflict problems. This suggests that the premises are actually having some effect here, despite what the protocol suggests.
Table 5.8
Classification of 'thinking aloud' protocols for Experiments 2 (n=16) and 3 (n=32) combined. Table (a) shows the percentage frequency of protocol classifications as a function of problem type. Table (b) shows the percentage of decisions favouring belief on each problem as a function of protocol classification (* indicates the logically correct response).

<table>
<thead>
<tr>
<th></th>
<th>C only</th>
<th>C to P</th>
<th>P to C</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>39</td>
<td>29</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>VU</td>
<td>34</td>
<td>28</td>
<td>34</td>
<td>3</td>
</tr>
<tr>
<td>IB</td>
<td>46</td>
<td>24</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>IU</td>
<td>41</td>
<td>33</td>
<td>17</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>C only</th>
<th>C to P</th>
<th>P to C</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>86*</td>
<td>100*</td>
<td>88*</td>
</tr>
<tr>
<td>VU</td>
<td>70</td>
<td>48</td>
<td>30</td>
</tr>
<tr>
<td>IB</td>
<td>73</td>
<td>78</td>
<td>54</td>
</tr>
<tr>
<td>IU</td>
<td>98*</td>
<td>97*</td>
<td>81*</td>
</tr>
</tbody>
</table>

Note:--
C only = Conclusion only
C to P = Conclusion to Premise(s)
P to C = Premise(s) to Conclusion
V = Valid
I = Invalid
B = Believable
U = Unbelievable
CONCLUSIONS AND GENERAL DISCUSSION

Experiments 1 to 3 have provided strong evidence for the existence of belief bias. The nature of the effect substantiates that of two previous studies which used a similar paradigm (Kaufman & Goldstein, 1967; Janis & Frick, 1943). On the syllogistic evaluation task, at least, there seems to be growing evidence to suggest that belief bias is a tendency to accept or reject conclusions according to their believability as opposed to their validity. However, this may be an oversimplification. Furthermore, this does not necessarily imply that belief bias would manifest itself in a similar way on other types of syllogistic task.

One factor which many of the earlier studies did not take into account was the illegal conversion of premises. Revlin et al (1978,1980) have argued that this process is responsible for the belief bias effect. Despite excluding illegal premise conversion, Revlin et al did, in fact, observe a belief bias effect. However they argue that this effect is relatively weak compared with that of logic. Experiments 1 to 3 challenge this claim. Controlling for premise conversion and other potentially confounding factors, each experiment showed a higher level of correct performance when logic and belief agreed than when they conflicted. Clearly, illegal conversion cannot explain the belief bias effect found here. Moreover, it would appear that belief bias is not as weak an effect as Revlin et al would suggest.

Experiments 1 to 3 not only found strong effects of both logic and belief, there was also evidence of an interaction
between the two (significant for the first two experiments and observed as a trend in the third). Such an interaction has also been reported by Kaufman & Goldstein (1967). From Tables 5.1, 5.3 and 5.6 we can see that belief exerts a greater effect when the conclusion given is believable than when it is unbelievable (i.e., logic appears stronger on problems with unbelievable conclusions). It also appears that belief bias is stronger on invalid problems.

The analysis of protocols has been useful in providing an indication of the basis of belief bias. However, it would be unwise to rely solely on such data. Henle (1962) based her characterisation of belief bias, and errors in general, on the interpretation of protocols. Inspection of protocols collected in Experiments 1 to 3 strongly suggests that Henle's method of analysis may be based on false assumptions. Henle claims that protocols show belief bias to result from rejection of the logical task. Indeed, in some cases this appears to be self-evident. (see Table 5.9(a) protocol 1). In other cases the classification of such explanations is less clear cut. Protocol 2 presents something of a dilemma. Is the subject saying that he does not believe the conclusion and this is why he rejects it, or is he saying that the (A) conclusion which he has (incorrectly) derived is inconsistent with the (O) conclusion presented? In this case, the protocol must simply be categorised as 'no reference to premises and no reference to irrelevant information'.

Other sources of error include the omission or distortion of premises and the misinterpretation of quantifiers. However, it is
clear that even when subjects omit or distort premises they can arrive at the correct conclusion. This is best illustrated with regard to conflict problems, on which logic and belief do not agree. Over Experiments 1 to 3 more than half of the correct responses are associated with either the omission or distortion of premises. The omission of one or more premises is a common occurrence in each experiment. Other studies have also noted this feature of protocols (Scribner, 1975; Cohen & Nagel, 1962).

Indeed, Scribner has shown that premises are often distorted and omitted in a simple syllogism recall task. It seems that subjects may easily forget precisely what they have paid attention to. Hence, we cannot always attribute judgemental errors to omissions in protocols. Unfortunately, when the premises are correctly restated this does not always lead to the correct decision (see protocol 3). Of course, such errors could be due to quantifier misinterpretation. However, this protocol provides no clues as to the nature of the possible misinterpretation. Other protocols suggest that the quantifier 'some' may frequently be misinterpreted (see protocol 4). This misinterpretation appears a number of times in protocols. It suggests that some subjects interpret "Some A are B" to mean "Some A are not B". This type of misinterpretation, along with the omission of premises, is cited by Scribner as a major source of difficulty when subjects are asked to restate syllogisms. However, in several cases, protocols do not tell us how the subject deviated from logical analysis because they are so brief (see protocols 5, 6 and 7 for examples). Unfortunately, if we reduced the protocol data to those which made sense this would be a wasteful and dubiously subjective
Example verbalisations illustrating the protocol scoring procedure. Each protocol is categorised according to two criteria: reference to both premises and reference to irrelevant information.

<table>
<thead>
<tr>
<th>SYLLOGISM PRESENTED &amp; PROTOCOL</th>
<th>PROTOCOL CLASSIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BP</td>
</tr>
</tbody>
</table>

1) No highly trained dogs are vicious,  
Some police dogs are vicious,  
Some police dogs are not highly trained  
(Valid-Unbelievable)

Some police dogs are vicious. But that's because they are trained to be so. In order to train a dog you must first of all; it must first of all trust you, and once you've gained that trust you can teach it to be vicious. But it must have, it must er, have some respect towards the master because otherwise I mean you wouldn't be able to do anything with the dog, it would just be a vicious animal and it would have no purpose. So in order for the police to use the dog it must be trained to be vicious, but it must also know when to obey a handler, and er. become used to some authority.  
(SUBJECT 32, EXPERIMENT 2).

2) No police dogs are vicious,  
Some highly trained dogs are vicious,  
Some police dogs are not highly trained.  
(Valid-Unbelievable)

This sentence at the end here - some police dogs are not highly trained - I'd say no. I'd say they are all highly trained.  
(SUBJECT 59, EXPERIMENT 2).
3) No rich people are hard workers,  
Some millionaires are hard workers,  
________________________  
Some rich people are not millionaires.  
(Invalid-Believable)  

All rich people can't be millionaires  
because no rich person works hard, whereas  
some millionaires do, therefore, some rich  
people must not be millionaires.  
(SUBJECT 23, EXPERIMENT 1).  

4) No vitamin tablets are inexpensive,  
Some nutritional things are inexpensive,  
________________________  
Some nutritional things are not vitamin  
tablets.  
(Valid-Believable)  

Some nutritional things are cheap,  
therefore, some must also be expensive. All  
vitamin tablets are expensive and they can  
also be the nutritional things which are  
expensive, which leaves the nutritional  
things which are cheap, and the nutritional  
things which are cheap are not vitamin  
tablets because they were the nutritional  
things which were expensive.  
(SUBJECT 38, EXPERIMENT 2)  

5) No rich people are hard workers,  
Some millionaires are hard workers,  
________________________  
Some millionaires are not rich  
(Valid-Unbelievable)  

Some millionaires are not rich I'd say no  
again.  
(SUBJECT 59, EXPERIMENT 2)  

6) No good swimmers are crewmen,  
Some deep sea divers are not crewmen,  
________________________  
Some deep sea divers are not good swimmers,  
(Valid-Unbelievable)  

No, this question says some and the passage  
says none do.  
(SUBJECT 2, EXPERIMENT 3)  

140
7) No judges are members of the committee, 
    Some well educated people are members of 
    the committee, 
    Some well educated people are not judges. 
    (Valid-Believable)

    No, I don't think it follows from the passage, 
    it has nothing to do with the passage 
    (SUBJECT 13, EXPERIMENT 3).

    Note:— REF TO BP = Reference to both premises 
     REF TO I = Reference to irrelevant information
exercise. The classification of protocols according to topic of reference, therefore, appears to be an appropriate method which is less dependent on the intuitions of the scorer. However, this method alone is of limited usefulness as Tables 5.2, 5.5 and 5.7 indicate, although there are a few illuminating trends here. The further classification of protocols according to the order in which syllogistic components are mentioned is more revealing. This enables us to trace the path taken through the problem, so long as the protocol is concurrent with problem analysis. Furthermore, this allows us to distinguish between protocols which would otherwise be categorised identically (as a simple reference to premises for example).

Table 5.9(b) gives example protocols which have been placed in one of the three categories: - 'premise(s) to conclusion', 'conclusion to premise(s)' and 'conclusion only'. Relating these classifications to the reasoning response given we find that most belief bias is associated with 'conclusion only' protocols. The lowest level of belief bias is associated with 'premise to conclusion' protocols, whereas 'conclusion to premise' protocols are associated with an intermediate level. With 'conclusion only' protocols, it seems that the premises were not considered relevant to solving the problems set. Subjects producing 'conclusion to premise' protocols appear to consider the premises relevant, but belief nevertheless influences their decision, albeit to a lesser extent overall. 'Premise to conclusion' protocols are relatively infrequent. Furthermore, they do not necessarily lead to the logically correct response, as we see from protocol 1 of table 5.9(b). Of course, there is more to
Examples of concurrent verbalisations falling into the three categories of: 'premise(s) to conclusion', 'conclusion to premise(s)' and 'conclusion only'.

**TYPE OF PROTOCOL AND SYLLOGISM PRESENTED**

**PREMISE TO CONCLUSION**

1) No healthy people are initial volunteers,  
Some astronauts are initial volunteers,  
Some astronauts are not healthy people.  
(Valid-Unbelievable)

It says that some astronauts are - some astronauts are initial volunteers which means that no - some of the astronauts which are the initial volunteers could not be healthy. So I don't think it's a sensible conclusion. No, the answer's no.  
(SUBJECT 1, EXPERIMENT 3)

2) No vicious dogs are highly trained,  
Some vicious dogs are police dogs,  
Some highly trained dogs are not police dogs.  
(Invalid-Believable)

No vicious dogs are highly trained, some vicious dogs are police dogs. You could say that some police dogs are not highly trained. But you can't say some highly trained dogs are not police dogs. So I disagree with that.  
(SUBJECT 15, EXPERIMENT 3)

**CONCLUSION TO PREMISE**

3) No good swimmers are crewmen,  
Some deep sea divers are crewmen  
Some deep sea divers are not good swimmers.  
(Valid-Unbelievable)

The statement some deep sea divers are not good swimmers is true, because it says some deep sea divers are crewmen, and then later it says no good swimmers are crewmen. This implies that good swimmers are not crewmen of which some deep sea divers are crewmen.  
(SUBJECT 8, EXPERIMENT 3)
4) No members of the committee are judges,
   Some members of the committee are well educated,
   Some well educated people are not judges.
   (Valid—Believable)

   Well, yes, the answer is yes; because it says that some members
   of the committee are well educated and there are no members of
   the committee that are judges, therefore, there are some educated
   people in the group yet there are no judges in the group.
   (SUBJECT 6, EXPERIMENT 3)

CONCLUSION ONLY

5) No priests are scientifically qualified
   Some religious people are scientifically qualified,
   Some priests are not religious people.
   (Invalid—Unbelievable)

   No, it doesn't really say anything about some priests are not
   religious people. It just says that science and religion don't go
   together.
   (SUBJECT 7, EXPERIMENT 3)

6) No initial volunteers are healthy people,
   Some initial volunteers are astronauts,
   Some healthy people are not astronauts.
   (Invalid—Believable)

   I would say to this statement of some healthy people are not
   astronauts, the answer is yes. The passage says that they have a
   large number of volunteers. Space flight as we know it is still
   in its very very early stages and for that reason you're going to
   have a very low percentage of astronauts existing anyway.
   Referring back to the point of the initial volunteers - numbering
   thousands I think, er, for that reason you are going to get some
   healthy people who have done another form of job and have not
   been an astronaut.
   (SUBJECT 10, EXPERIMENT 3)
reasoning than just focusing on the logically relevant information.

So far protocol analysis has associated different modes of reasoning with different levels of belief bias. However protocols do not tell us why people adopt a particular mode. It seems reasonably safe to assume that 'conclusion only' protocols largely reflect a rejection of the logical task. 'Premise to conclusion' protocols may result from a straightforward attempt at logical analysis. Indeed most theories of syllogistic reasoning assume that reasoning proceeds in this manner. However, this does not appear to be a dominant approach when the given conclusion has obvious truth status. The 'conclusion to premise' protocols are difficult to explain.

In some ways this type of protocol is the most interesting, since it suggests that belief bias need not simply reflect a rejection of the logical task nor direct evaluation of the conclusion. Here we have evidence that some subjects, who consider the premises, are influenced in some way by the believability of the conclusion presented. There are various ways in which logic and belief could exert an influence here. Perhaps the two factors receive different weightings depending on the validity or believability of a given syllogism.

In conclusion we can say that Experiments 1 to 3 have made considerable headway in establishing and characterising the belief bias effect. Since belief bias occurs on syllogisms with convertible premise pairs, it cannot be attributed to their illegal conversion. Furthermore, we know that belief bias is strongly associated with those protocols which focus primarily on...
exclusively on the conclusion. This might suggest that if subjects could be encouraged to consider the premises and reason in a 'forward' premise to conclusion fashion, the belief bias effect should be considerably reduced.
Chapter 6

Belief bias and the syllogistic construction task.

General Introduction 148

Experiment 4
Introduction 150
Method & Design 153
Categorisation of responses 157
Results & Discussion 157

Experiment 5
Introduction 168
Method & Design 169
Results & Discussion 173

General Discussion 184
Conclusions & Implications 190
INTRODUCTION

The experiments described in Chapter 5 suggested that the majority of belief biases could be associated with a direct belief-based evaluation of the conclusion or another mode of (backward) reasoning. Evaluation of a given conclusion merely requires the subject to test the logical relation between the premises and the conclusion given. The subject may thus work from premises to conclusion or alternatively he may work 'backwards' from conclusion to premises. Particularly when the problem content is thematic, conclusion-based strategies or heuristics could be further encouraged, hence compounding the problem. Whilst some psychologists (eg Brickson, 1975) are willing to assume task equivalence, others (eg Wason & Johnson-Laird, 1972; Johnson-Laird & Steedman 1978) argue that there may be fundamental differences between different types of syllogistic task. Indeed Aristotle, the inventor of syllogisms, commented on this issue, stating that the arguments that are best understood are those which the audience follows, but for which the conclusion is not immediately obvious (Linder & Worchel, 1970). It is, therefore, puzzling to say the least to find that almost all studies of belief bias have been based exclusively on the conclusion evaluation or response selection paradigm. The findings so far raise two questions in particular: a) What would happen if the invitation to process information in a backward fashion were withdrawn? and b) to what extent is belief bias reliant on
problem presentation? Both questions may be addressed by replacing the syllogistic evaluation task with one requiring conclusion construction.

The two experiments described in this chapter examine the effect of belief bias using the syllogistic construction task. Experiment 4 seeks to test the generalisability of the belief bias effects found in Experiments 1 to 3. This experiment is, therefore, essentially exploratory. Experiment 5 provides a more detailed follow-up study.
EXPERIMENT 4

The problems used in this experiment differed from those used in Experiments 1 to 3 in three major ways:

(i) Premise pairs only were presented.
(ii) Premise pairs were not embedded in any form of prose passage.
(iii) Indeterminate premise pairs were included; these took the place of the invalid problems used in Experiments 1 to 3.

There are two reasons why premises were not embedded in prose passages. Firstly, the syllogistic construction task calls for detailed instructions concerning the exact logical form of the conclusion produced. The embedding of syllogistic premises in prose was, therefore, likely to contribute to the complexity of the task and encourage the illegal combination of irrelevant material within the conclusion produced. Any variation from the conventionally accepted conclusion was likely to produce scoring difficulties and reduce the number of acceptable conclusions.

A second problem is that the scenario may artificially induce the use of a belief-based strategy. Although this possibility was ruled out in the case of the evaluation task, it cannot be assumed, using alternative paradigms, that subjects will not be led to expect certain conclusions as a result of the accompanying prose passage. The embedding of syllogistic premises was, therefore, rejected as a presentation format, in order to rule out the possible intrusion of artificially induced belief-based strategies.
Indeterminate premise pairs were incorporated to broaden the scope of the problems used. In Experiments 1, 2 and 3, invalid problems were created by combining determinate premise pairs with an invalid conclusion. Since this method of invalidation is not possible with the construction paradigm, it was necessary to produce indeterminate premise pairs for inclusion. This was achieved by combining E and O premises. Indeterminate problems were based on existing, determinate, premise pairs and were produced by replacing the I premise with an O premise. The transformation is illustrated by the following four, example, problems. The four problems represent each of the four problem types used in Experiment 4.

**Determinate premises (L+), yielding believable (B+) conclusion**

No vicious dogs are police dogs,  
Some vicious dogs are highly trained dogs,  
Therefore:  
Some highly trained dogs are not police dogs.

**Indeterminate premises (L-), based on above**

No vicious dogs are police dogs,  
Some vicious dogs are not highly trained dogs,  
Therefore:  
No valid conclusion
Determinate premises (L+), yielding an unbelievable (B-) conclusion
No vicious dogs are highly trained dogs,
Some vicious dogs are police dogs,
Therefore:
Some police dogs are not highly trained dogs.

Indeterminate premises (L-), based on above
No vicious dogs are highly trained dogs,
Some vicious dogs are not police dogs,
Therefore:
No valid conclusion.

It should be noted that both determinate and indeterminate premise pairs produce an atmosphere which favours an O conclusion. The superficial similarity of EI and EO premise pairs is intended to reduce the likelihood of problem differentiation on the basis of 'surface' features. Since determinate and indeterminate premise pairs cannot be distinguished on the basis of atmosphere, any differential effects cannot be confounded with differential atmosphere effects. However, we know from empirical evidence that differences exist in response profiles to EI and EO premise pairs. (see response frequencies reported by Johnson-Laird & Steedman, 1978).
METHOD

Half the problems consisted of determinate premise pairs, and the remaining half indeterminate premise pairs. Half the determinate premise pairs were BI and half were IE. Half the indeterminate premise pairs were BO and half OE.

All valid conclusions accorded with atmosphere. In half of these cases the valid conclusion was believable and in the remaining half it was unbelievable.

All problems were structured on the basis of a figure 3 syllogism:

\[
B - A \\
B - C
\]

For this figure, Johnson-Laird & Steedman report no reliable bias towards one or other direction (A - C or C - A) of conclusion.

The eight problem contents presented in this experiment were identical to those used in Experiment 3.

Design

Each subject received eight problems, four determinate, and four indeterminate premise pairs. The four determinate premise pairs yielded two Valid-Believable and two Valid-Unbelievable conclusions in all. They could each yield an Invalid-Believable conclusion.

The design was balanced such that each problem type occurred
with each problem content an equal number of times. Presentation order was randomised. Problems were presented in booklet form. Subjects were run in groups of four.

Subjects

Thirty-two undergraduates at Plymouth Polytechnic acted as paid volunteers. Subjects had no previous experience of this task nor any training in logic.

Procedure

Task and instructions

Instructions and problem booklets were photocopied. Instructions were retained by subjects for reference during the experiment.

The subjects' task was to decide what conclusion, if any, could be deduced from the information given in the premises of each problem. A space was provided below the premises in which each conclusion was to be written.

Detailed instructions were given concerning the format of conclusions produced. The logical meaning of the quantifier "some" was also elaborated. This was included for two reasons. Firstly, the syllogistic construction task experiments on which much of this design was based (i.e. Johnson-Laird & Steedman, 1978) included instruction on the interpretation of "some". It is unclear at this stage exactly how subjects interpret this quantifier (see Chapter 2, Section 2). This part of their procedure was, therefore, incorporated in order to make direct comparisons between the two studies more legitimate, since
differences in response profiles could be caused by misinterpretation of this quantifier. Secondly, since all subjects received both I and O premises (in combination with E premises) it seemed quite likely that the two appearances of the word "some" may confuse the subject and hence complicate his interpretation of premises. Explanation of the quantifier "some" would, therefore, reduce any confusion over premise interpretation.

The exact instructions given to subjects were as follows:

This experiment is designed to find out how people solve logical problems. In the booklet which you have been given there are 8 logical reasoning problems. Your task is to decide what conclusion, if any, can be derived from the information which you have been given. The information takes the form of two statements (premises) which can be expressed symbolically as follows:

ALL B ARE A,
SOME C ARE B.

As you can see, the two premises tell us something about the relationship between three terms: A, B and C. The term B never appears in the conclusion; since the conclusion is a statement about the relationship between A and C, or vice versa. The conclusion to the above example is, therefore, SOME C ARE A.

Since this is a problem requiring logical analysis, you should interpret the word 'SOME' in its strictly logical
sense; meaning AT LEAST ONE AND POSSIBLY ALL. So the statement "SOME B ARE C" does not necessarily also mean that SOME B ARE NOT C.

In the booklet you will find 8 different logical problems. They are the same type of problem as the example problem which we have shown above, however, the terms used will not be letters of the alphabet, but real words instead. Your task is to write down, in the booklet provided, the conclusion which you think logically follows from the two premises in the space provided below each problem. Please do not simply restate one or more of the premises as your conclusion.

You are reminded that you must produce a conclusion based on the information given in the two premises - and this information only. 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 specific conclusion logically follows from the information given you should write it down; there may or may not always be a definite conclusion derivable from each problem. If you think there is no definite conclusion to a problem, then you should simply write "NO VALID CONCLUSION" below the premises.

Please take your time and be certain that you have the logically correct conclusion 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 in front of you in case
you need to refer to them later on.

REMEMBER, YOUR CONCLUSIONS SHOULD BE BASED SOLELY UPON WHAT CAN BE DEDUCED WITH ABSOLUTE CERTAINTY FROM THE TWO PREMISES — AND THIS INFORMATION ONLY

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 aid you in this task.

Categorisation of responses

Conclusions which deviated from the usual logical format were reconstructed where possible. The method of reconstruction was decided upon by a panel of three judges, and thereafter implemented by a single judge. If any conclusion could not be reconstructed either because it contained additional (logically irrelevant) material or the conclusion was simply a restatement of the premises it was categorised as "unclassifiable".

RESULTS AND DISCUSSION

The great majority of classifiable conclusions were written in the usual logical form. However, a small number of logically convoluted conclusions were produced and consequently reconstructed. For example:
"All police dogs are not highly trained" was reconstructed as "No police dogs are highly trained".
"Not all highly trained dogs are police dogs" was reconstructed as "Some highly trained dogs are not police dogs".
"Highly trained dogs that are not vicious can be police dogs" was reconstructed as "Some highly trained dogs are police dogs". This type of response has also been noted by Wason & Johnson-Laird (1972) and Johnson-Laird & Steedman (1978), and is referred to as "partial digestion of the middle term". The conclusion itself is logically impeccable, however, it is not the most economical way of expressing the conclusion.

The terms "most" and "not always" were translated as "some" and "some are not" respectively.

The manner in which conclusions are translated is a debatable point. An important consideration concerns whether the strict logical meaning of the conclusion should be retained, or whether linguistic conventions should also be taken into account. From the review of the interpretation and use of quantifiers it is clear that the two types of usage may deviate. Wherever possible, therefore, in reconstructing conclusions an attempt was made to bear in mind the original intention of the subject.

The number of conclusions classified as believable are shown in table 6.1. Tables 6.1 (a) and (b) show data for determinate and indeterminate premise pairs respectively. This provides an indication of the degree and nature of any belief bias present on this task. The number of unbelievable conclusions which fall into the four moods can be ascertained by subtracting the number of believable conclusions from the total number for that category.

Excluding unclassifiable conclusions, the overall percentage of believable conclusions was 42%. For determinate premise pairs yielding a Valid-Believable (O) conclusion 53% of conclusions produced were believable. This figure dropped
Table 6.1(a)

Frequency of responses for determinate premise pairs, divided according to mood and believability of conclusion produced. Experiment 4 (n=32)

**Determinate**

<table>
<thead>
<tr>
<th>Mood</th>
<th>Valid-Believable (O) conclusion (L+B+)</th>
<th>Valid-Unbelievable (O) conclusion (L+B-)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of conclusions</td>
<td>Number of believable conclusions</td>
</tr>
<tr>
<td>A</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>E</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>O</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Non-propositional</td>
<td>12</td>
<td>--</td>
</tr>
<tr>
<td>Unclassifiable conclusion</td>
<td>17</td>
<td>--</td>
</tr>
<tr>
<td>TOTAL</td>
<td>64</td>
<td>25</td>
</tr>
</tbody>
</table>
Table 6.1(b)
Frequency of responses for indeterminate premise pairs, divided according to mood and believability of conclusion produced.
Experiment 4 (n=32).

<table>
<thead>
<tr>
<th>Mood</th>
<th>Based on L+B+</th>
<th>Based on L+B-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of conclusions</td>
<td>Number of believable conclusions</td>
</tr>
<tr>
<td>A</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>O</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Non-</td>
<td>18</td>
<td>--</td>
</tr>
<tr>
<td>propositional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unclassifiable</td>
<td>20</td>
<td>--</td>
</tr>
<tr>
<td>conclusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>64</td>
<td>24</td>
</tr>
</tbody>
</table>
dramatically to 18% for determinate premise pairs, yielding a Valid-Unbelievable (O) conclusion. This fall is mainly due to the large number of correct (O) conclusions produced. For indeterminate premise pairs based on determinates yielding Valid-Believable and Unbelievable conclusions the percentage categorised as believable was 55% and 45% respectively. Overall, more than half (58%) of conclusions produced were either unbelievable or non-propositional.

It could be argued that the relatively high percentage of unclassifiable conclusions conceals evidence of belief bias. However, if we examine the pattern of unclassifiable conclusions there are only a minority resulting from the inclusion of external, irrelevant material. The majority of unclassifiable conclusions resulted from the inclusion of the B (middle) term with either the A or C term. Although the majority of conclusions may not be termed unbelievable, they are by no means any more believable than the A-C or C-A conclusions available. It is, therefore, concluded that the majority of unclassifiable conclusions were not the result of belief bias.

There seems to be no bias against producing unbelievable or non-propositional conclusions on this task. There is also little evidence of the, non-logical, atmosphere effect (Woodworth & Sells, 1935). Atmosphere predicts a predominance of O conclusions. Inspection of Tables 6.1 (a and b) shows a wide variety of conclusions which cover all moods. Indeterminate problems, in fact, show a large number of A and I conclusions, with a predominance of the logically correct non-propositional conclusion. For determinate problems the O conclusion is also
predicted by logic so it cannot be viewed as an exclusive measure of atmosphere bias. However, the overall pattern of responses also departs from that predicted by other theories of syllogistic reasoning. In the construction task study reported by Johnson-Laird & Steedman, the pattern observed did not accord with atmosphere, but did not show the diversity observed here. For EI, IE, EO and OE premise pairs in the B-A B-C figure, only three types of conclusion are recorded: E, O and the non-propositional conclusion. For the determinate premise pairs, the dominant response is the O conclusion, closely followed by the non-propositional conclusion. A relatively small percentage of E conclusions are also reported. For the indeterminate premise pairs, by far the most dominant response is the non-propositional conclusion. If we look at overall accuracy, it is clear that performance drops considerably below that observed by Johnson-Laird & Steedman. In Experiment 4 26% of conclusions were correct for determinate premise pairs. For indeterminate premise pairs 34% of conclusions were correct. For the same figure / mood combinations Johnson-Laird & Steedman report approximately 52% and 72% correct for determinate and indeterminate premise pairs respectively (figures totalled over both tests). Clearly, the overall accuracy in Experiment 4 falls to roughly half this level.

Since the data obtained in Experiment 4 do not follow the pattern shown by existing research, it is necessary to check those aspects of the data for which assumptions were made in designing the experiment. On the basis of data provided by Johnson-Laird & Steedman it was predicted that no (figure based)
directionality effects would be observed in the conclusions produced. Clearly, it is necessary to identify any such unpredicted effects which may have contributed to the overall lack of belief bias. Tables 6.2(a) and (b) show the frequency of conclusions divided according to their directionality, i.e., whether they run from A-C or C-A.

Both tables present conclusions, regardless of their believability. The tables differ inasmuch as Table 6.2(a) shows data for determinate premise pairs, and 6.2(b) shows data for indeterminate premise pairs. In each table, the term A refers to the first term (other than the middle or linking term) to appear in the premises; the term C refers to the final term to appear thus:

\[
\begin{align*}
B & \rightarrow A \\
B & \rightarrow C
\end{align*}
\]

Therefore: ?

In response to any premise pair, two types of propositional conclusion may be produced for each mood. Conclusions may run from A to C or, conversely, they may run from C to A. For figure 3 syllogisms, Johnson-Laird & Steedman report no bias towards either direction of conclusion.

In the case of the present experiment, Tables 6.2(a) and 6.2(b) indicate a very slight tendency to produce A to C as opposed to C to A conclusions. However, this tendency does not characterise performance on all problem types. Indeed, directionality preferences appear to vary considerably, according...
Table 6.2(a)

Frequency of conclusions to determinate premise pairs, divided according to problem type, premise order, mood and direction of conclusion. Experiment 4 (n=32).

<table>
<thead>
<tr>
<th>Determinate</th>
<th>Valid-Believable (O) conclusion (L+B+)</th>
<th>Valid-Unbelievable (O) conclusion (L+B-)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EI</td>
<td>IE</td>
</tr>
<tr>
<td>Conclusion</td>
<td>A-C</td>
<td>C-A</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>O</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 6.2(b)

Frequency of conclusions to indeterminate premise pairs, divided according to determinate premise pairs from which each was derived, premise order, mood and direction of conclusion. Experiment 4 (n=32).

<table>
<thead>
<tr>
<th>Indeterminate</th>
<th>Based on L+B+</th>
<th>Based on L+B-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EI</td>
<td>IE</td>
</tr>
<tr>
<td>Conclusion</td>
<td>A-C</td>
<td>C-A</td>
</tr>
<tr>
<td>A</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>O</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>
to problem type and order of premises. This pattern is clearly most evident in Table 6.2(b).

With specific reference to Table 6.2(b), for EO premises, based on Determinate-Believable problems, there is a definite preference for A to C conclusions. For the same problem type, but with the OE premise combination, the opposite bias exists, and C to A conclusions are favoured. Turning to EO and OE premise combinations, based on Determinate-Unbelievable problems, an alternative pattern emerges. In this case, there is a bias towards C to A and A to C conclusions respectively.

For Table 6.2(a), which presents data for determinate problems, trends are less clear. Directionality preferences are marked for Determinate-Unbelievable problems only.

Examination of the actual problem composition, on which Tables 6.2(a) and 6.2(b) are based suggests an interesting, if somewhat puzzling, account of the observed trends. Since only Table 6.2(b) shows clear evidence of the bias, for simplicity, only problems on which this table is based will be considered in detail. Note that this table represents indeterminate problems which are closely based on determinate problems and differ in mood only. The basis of trends observed on indeterminate problems also, therefore, characterises those observed on determinate problems.

The following four example problems constitute the problems represented by columns one to four in Table 6.2(b). They are numbered accordingly:
(1) No vicious dogs are police dogs,

Some vicious dogs are not highly trained dogs,

Therefore:

(2) Some vicious dogs are not highly trained dogs,

No vicious dogs are police dogs,

Therefore:

(3) No vicious dogs are highly trained dogs,

Some vicious dogs are not police dogs,

Therefore:

(4) Some vicious dogs are not police dogs,

No vicious dogs are highly trained dogs,

Therefore:

Working through each problem individually, problem (1) represents column one. In this column there is a clear tendency for conclusions to run from A to C (from 'police dogs' to 'highly trained dogs'). Problem (2) represents column two, where there is a bias towards C to A conclusions (from 'police dogs' to 'highly trained dogs'). Problem (3) represents column three in which there is a tendency to form C to A conclusions (from 'police dogs' to 'highly trained dogs'). Problem (4) represents column four, where there is a bias towards A to C conclusions ('police dogs' to 'highly trained dogs'). Since the linguistic and semantic characteristics of this set of example problems are identical to those of all problem contents used in this
experiment, the observed directionality preferences can be characterised as either one of two biases. A conclusion which runs from 'police dogs' to 'highly trained dogs' can be seen to run from subset to superset, since there are more highly trained dogs than police dogs in the world. Alternatively the same conclusion can be seen to run from noun phrase to adjectival noun phrase.

The proposed bias(es) also characterise, not only all data pertaining to indeterminate problems, but also half of the data for determinate (ie unbelievable) problems shown in Table 6.2(a). The reason for this tendency is unclear. Inspection of response patterns to each of the eight problem contents indicates that this bias is not attributable to any peculiarity in responding to any particular problem content. The bias can, therefore, be viewed as reflecting a genuine, overall, tendency.

In order to clarify the interpretation of results it is first necessary to replicate the present findings. Secondly, to establish whether there are any thematic effects it is also necessary to compare performance on logically equivalent abstract and thematic tasks. Finally, if there is a directionality bias in conclusion production we must be able to experimentally differentiate between a bias to place sub before superset and a bias to place noun before adjectival noun phrase.
EXPERIMENT 5

Experiment 4 found no evidence of a belief bias effect. However, response patterns within the conclusions produced suggested a hitherto undetected directionality bias. Experiment 5 was designed as a more detailed follow-up study which had three basic aims. The first was to replicate the findings of Experiment 4. Hence, one group of subjects received a set of problems which included those given to subjects in Experiment 4. The second aim was to further investigate the purported directionality bias. In order to distinguish between a bias to place sub before superset and a bias to place noun phrase before adjectival noun phrase, another collection of thematic problems was created. This set of problems was termed the 'Thematic Comparison' set and contained conclusions which had no noun/adjectival noun phrase distinction; they either contained noun phrases or adjectival noun phrases - never both. Because of the very nature of the conclusions used, the Thematic Comparison set had to retain the sub/superset distinction. Hence, if the directionality bias was present in the conclusions produced to Thematic Comparison problems it could then be seen as a preference to place sub before superset. The third aim of Experiment 5 was to compare performance on thematic problems with that on abstract problems of identical logical form. Belief bias is not always exerted in a straightforward fashion (see Revlin et al, 1980; Małczycki, 1978). Hence, in this type of task, beliefs may be affecting performance in a more subtle manner than might be predicted. Comparisons between performance on thematic and abstract control problems would, therefore aid the detection of any content-based peculiarities in
responding.

**METHOD**

The problems used in Experiment 5 were divided into three main sets as follows:

**Set One: Thematic Replication**

This set of problems was identical to those used in Experiment 4. Propositions in this set confounded category size and linguistic status, such that the subset (e.g., police dogs) was also a noun phrase and the superset (e.g., highly trained dogs) was also an adjectival noun phrase.

**Set Two: Thematic Comparison**

This set of problems contained propositional conclusions whose terms were distinguishable as sub and superset. The linguistic form of end terms within any given problem remained constant, i.e., all conclusions contained either two noun phrases or two adjectival noun phrases. The two forms were never mixed within the same conclusion. Examples of the two types of propositional form are as follows:
Noun phrase to noun phrase
No novels are books.
No cars are vehicles.
No chocolates are sweets.
etc.

Adjectival noun phrase to adjectival noun phrase:
No Victorian furniture is antique furniture.
No serious accidents are fatal accidents.
No rich actors are wealthy people.
etc.

Set Three: Abstract Control
This set contained abstract premise pairs whose terms were letters of the alphabet. The logical structure of abstract problems was identical to that of thematic problems.

(NOTE: The moods and structural composition of these thematic problems and the basic make-up of problem sets was identical to the set outlined in Experiment 4. A number of extra problems were also added to increase scope, however.)

Additional problems
In Experiment 4 all E conclusions were necessarily unbelievable. All I conclusions were believable. This restricted the type of permissible conclusions. This state of affairs is undesirable, and so, in Experiment 5, problems which allowed
believable E and unbelievable I conclusions were also included. This broadened the scope of possible E and I conclusions available to subjects. This meant that a believable or unbelievable conclusion was possible in each mood at some time during the experiment.

The additional problems were constructed by combining EA, AE and EE premises. The structure of these premise pairs was based on the figure 2 syllogism: A - B/C - B. This alteration in figure was necessary in order to allow valid conclusions to accord with atmosphere predictions. No figure 3 syllogisms yield valid E conclusions, which accord with atmosphere. The two alternative figures (1 and 4) are unsuitable for use in this case, since strong directionality biases have been observed for these figures (see Johnson-Laird & Steedman, 1978).

As for other problems sets, half of the additional problems were determinate and half indeterminate. Half of the determinate problems were EA and half were AE premise pairs. All the indeterminate problems were EE premise pairs.

Four (two determinate and two indeterminate) problems from this group were assigned to each of the three problem sets (Thematic Replication', 'Thematic Comparison' and 'Abstract Control').

The believability of all valid conclusions, which arose from the construction of additional problem material, was ascertained by questionnaire as in earlier experiments (see Appendix IV). No subjects who took part in the questionnaire study were allowed to take part in the experiment proper.
Design

Subjects were randomly allocated to one of three groups which were differentiated according to the set of problems given. Each subject received ten problems, each drawn from a given problem set. Two of the ten problems consisted of items drawn from the additional pool of problems; which allowed believable E and unbelievable I conclusions. Half of the problems were determinate (half believable and half unbelievable) and half were matched indeterminate problems.

Combination of problem type and content was balanced for each set. As in Experiment 4, problems were presented in booklet form and presentation order was randomised. Subjects were run in groups of four.

Subjects

Forty-eight undergraduates at Plymouth Polytechnic acted as paid volunteers. Subjects had no previous experience of this task nor any training in logic.

Procedure

Task and Instructions

The task and instructions for groups receiving thematic problems were identical to those of Experiment 4. The instructions which accompanied abstract problems differed from those used for thematic problems with respect to the content of the example problem and propositions given. The relevant section
is as follows:

ALL BEE KEEPERS ARE ARTISTS,
SOME CHEMISTS ARE BEE KEEPERS,

"As you can see, the two premises tell us something about the relationship between BEE KEEPERS, ARTISTS AND CHEMISTS. The term "BEE KEEPERS" never appears in the conclusion; since the conclusion is a statement about the relationship between "ARTISTS" and "CHEMISTS" or vice versa. The conclusion to the above example is therefore: SOME CHEMISTS ARE ARTISTS.

Since this is a problem requiring logical analysis, you should interpret the word 'SOME' in its strictly logical sense; meaning AT LEAST ONE AND POSSIBLY ALL. Thus, "SOME DOCTORS ARE GOLFERS" does not necessarily mean "SOME DOCTORS ARE NOT GOLFERS".

Categorisation of responses
Responses were categorised in the same manner as described for Experiment 4.

RESULTS AND DISCUSSION

Experiment 5 set out to answer three main questions. Firstly, can the results of Experiment 4 be replicated? Secondly, what is the cause of the suspected linguistic directionality bias and is it a significant effect? Finally, is there any other thematic effect which may have gone unnoticed in Experiment 4? The aims of Experiment 5 are dealt with separately in this
Is there a belief bias effect?

Tables 6.3(a and b) present the frequency of believable and unbelievable conclusions produced over both thematic groups, for determinate and indeterminate problems respectively.

Clearly, both unbelievable and non-propositional conclusions outnumber believable conclusions. Overall, 29% of categorisable conclusions are believable, whereas 35% are unbelievable. However, this does not characterise the pattern of conclusions for all problem types. For indeterminate problems there are more believable than unbelievable conclusions (as was the case in Experiment 4), but the frequency of believable conclusions is far outweighed by that of non-propositional conclusions. Examination of Table 6.3(a) suggests an interesting interaction on determinate problems, which might be predicted on the basis of logic. On Valid-Unbelievable problems there appears to be a strong tendency to produce unbelievable conclusions. For Valid-Believable problems, a slight opposite tendency is apparent, more believable than unbelievable conclusions seem to be produced. If this pattern is indeed due to the influence of logic, it should follow that substantially more correct responses were given to the Valid-Unbelievable than to the Valid-Believable problems (notice that this runs counter to any belief bias predictions). For both types of determinate premise pair, the percentage correct (disregarding unclassifiable conclusions) is alarmingly low. For the Valid-Believable problem type, 15% of conclusions are correct; for the Valid-Unbelievable problem type
### Table 6.3(a)

Frequency of conclusions produced for determinate premise pairs, divided according to mood and believability, pooled over both thematic groups. Experiment 5 (n=32).

<table>
<thead>
<tr>
<th>Determinate</th>
<th>Valid-Believable (O) conclusion (L+B+)</th>
<th>Valid-Unbelievable (O) conclusion (L+B-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of conclusions</td>
<td>Number of believable conclusions</td>
<td>Number of conclusions</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>O</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Non-propositional</td>
<td>17</td>
<td>--</td>
</tr>
<tr>
<td>Unclassifiable conclusion</td>
<td>16</td>
<td>--</td>
</tr>
<tr>
<td>TOTAL</td>
<td>64</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 6.3(b)

Frequency of conclusions produced for indeterminate premise pairs, divided according to mood and believability, pooled over both thematic groups. Experiment 5 (n=32).

<table>
<thead>
<tr>
<th></th>
<th>Based on L+B+</th>
<th>Based on L+B-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of conclusions</td>
<td>Number of believable conclusions</td>
<td>Number of conclusions</td>
</tr>
<tr>
<td>A</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>O</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Non-propositional</td>
<td>26</td>
<td>--</td>
</tr>
<tr>
<td>Unclassifiable conclusions</td>
<td>19</td>
<td>--</td>
</tr>
<tr>
<td>TOTAL</td>
<td>64</td>
<td>14</td>
</tr>
</tbody>
</table>

Note: The pattern of classifiable conclusions is similar for both types of thematic content. However, the frequency of unclassifiable conclusions is higher for the thematic comparison than the thematic replication group. The reason for this difference is unclear.
a comparably low 21% are correct. Clearly, the overwhelming preponderance of unbelievable conclusions produced in response to the latter problem type cannot be attributed to validity effects. An examination of the distribution of responses over all moods for the two determinate premise pairs suggests that the observed imbalance in the believability of responses is due to a preponderance of E conclusions to the Valid-Unbelievable compared with the Valid-Believable problem type. Exactly why this differential response pattern should exist is not immediately obvious. However, inspection of response distributions for the abstract problems helps to solve the problem. Table 6.4 shows the distribution of responses for the Abstract Control group.

An interesting picture emerges here. For abstract, determinate premise pairs, the E conclusion constitutes the dominant response. This suggest that the response pattern observed on the Valid-Unbelievable problem type may not be an artifact of thematic characteristics. The dominance of E conclusions may be due to problem structure as opposed to content. Indeed, responses to the Valid-Believable problem show a slight bias towards this conclusion, albeit relatively small.

Unlike Experiment 4, Experiment 5 provides the opportunity to compare performance on problems of the same logical status (determinate or indeterminate), with a differing figure/mood combination. In Experiment 5, additional problems were constructed to allow some believable E and unbelievable I conclusions. This meant that throughout the entire problem set, received by each subject, no mood of conclusion was exclusively associated with one type of belief (ie positive or negative).
distribution of responses to these (figure 2) problems followed much the same trend for both thematic and abstract content. Clearly, the believability of conclusions had no effect on responding. Although there are more E than I conclusions produced for the thematic group (which, in fact means there are more believable than unbelievable) this is also the case for the Abstract Control group. We can see this pattern in Table 6.5.

It is evident from the examination of responses so far, that belief bias has once again disappeared on the syllogistic construction task. Problem content can have various effects, of which belief bias is only one. The distribution of responses over the five conclusion types could be another way in which problem content finds expression. As we have already seen in Table 6.5, this is not the case for the figure 2 problems, but we do not know for sure whether this is also the case for the figure 3 problems used. In order to test whether the distribution of responses over the five conclusion types (disregarding direction) differed significantly between the three main problem sets, the proportion of total responses falling into each of the five response categories was calculated for each subject (disregarding unclassifiable conclusions). Kruskall Wallis tests, comparing each of the three groups for each conclusion type separately, revealed only one significant difference between groups; which was for I conclusions (p < 0.05, two-tailed). A multiple comparison test (using the formula due to Dunn (1964)) indicated a significant difference between the Abstract and Thematic Comparison group for this conclusion type. The fact that such a difference exists between the Abstract and only one of the two
Table 6.4

Frequency of responses for the Abstract Control group, Figure 3 problems. Experiment 5 (n=16)

<table>
<thead>
<tr>
<th>Mood of conclusion</th>
<th>Determinate</th>
<th>Indeterminate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direction of Conclusion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A-C</td>
<td>C-A</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>I</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>O</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Non-propositional</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.5

Frequency of classifiable conclusions to determinate (EA, AE) and indeterminate (EE) premise pairs which allow believable E and unbelievable I conclusions. Alongside are the corresponding conclusions produced to the matched Abstract Control problems. Experiment 5 (n=48), both thematic groups combined.

<table>
<thead>
<tr>
<th>Conclusion produced</th>
<th>Thematic problems</th>
<th>Abstract control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>I</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>O</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Non-propositional</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

D = Determinate (EA, AE premise pairs)
I = Indeterminate (EE premise pairs)
Thematic groups is puzzling. The inexplicability of this finding suggests that this may have been due to a Type 1 error. Overall, it is clear from this analysis that problem content did not fundamentally affect the production of conclusions in terms of mood.

So far, two general effects of content can be ruled out. Subjects do not seem to be influenced by the believability of conclusions; neither does the content of the problem seem to influence the relative frequency of the various (five) conclusion types. One very specific, thematic-based, effect could still occur, however. A preference may exist to place certain terms before others when producing a conclusion.

Is there a directionality bias in conclusion production?

All categorisable conclusions were classified as one of two types:
(i) those which ran from subset (or noun phrase) to superset (or adjectival noun phrase).
(ii) those which ran in the opposite direction.

Tables 6.6 (a and b) set out the frequency of conclusions which do and do not conform to the directionality bias (category (i) and (ii) respectively).

A Wilcoxon matched pairs test was carried out on the two classifications (for and against). For each group of subjects, comparing the total frequency of conclusions falling into each category for each subject. A significant difference between the two conclusion categories was found for both thematic groups (p < 0.05, one-tailed, in both cases). This indicates that there
is indeed a significant thematic directionality effect on conclusion production when this particular type of thematic content is used. In order to detect any differences in directionality scores between the two groups, an overall directionality score was calculated for each subject by subtracting the total number of conclusions which went against the bias from the total of those which conformed to it. A Mann-Whitney test comparing scores across thematic groups showed no significant differences between them ($p > 0.05$). The fact that the effect is similar in both thematic groups tells us something about its nature. Recall that the two thematic sets of problems differed on one dimension only; the linguistic relationship between the two constituent terms of propositional conclusions. To recap; the Thematic Replication group had conclusions containing a noun phrase and an adjectival noun phrase. The Thematic Comparison group had conclusions containing either two noun phrases or two adjectival noun phrases - the two types of phrase were on no occasion present in the same conclusion. For both thematic groups each propositional conclusion had a semantically defined subset and superset term. Since the directionality bias was present in both groups, we have to conclude that the bias is a semantic one, based on set size. Specifically, it is a preference to place subset before superset.

Inspection of tables 6.6 (a and b) indicates that, in overall terms, the bias is present on all problems, except the Valid-Believable type. Note that the conclusion prescribed by logic runs against the proposed bias, in this case. This appears to have made a major contribution in swaying the overall
Table 6.6(a)

Frequency of classifiable conclusions in each of the four moods which do and do not conform to the sub-superset directionality bias. This table gives figures for determinate premise pairs with both thematic groups combined. Experiment 5 (n=32).

<table>
<thead>
<tr>
<th>Mood of conclusion</th>
<th>Determinate</th>
<th>Valid-Believable</th>
<th>Valid-Unbelievable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0 conclusion (L+B+)</td>
<td>0 conclusion (L+B-)</td>
</tr>
<tr>
<td></td>
<td>For</td>
<td>Against</td>
<td>For</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>O</td>
<td>0</td>
<td>7*</td>
<td>11*</td>
</tr>
<tr>
<td>Non-propositional</td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Logical conclusion is marked by an asterisk.
Table 6.6(b)

Frequency of classifiable conclusions in each of the four moods which do and do not conform to the sub-superset directionality bias. This table gives figures for indeterminate premise pairs with both thematic groups combined. Experiment 5 (n=32).

<table>
<thead>
<tr>
<th>Mood of conclusion</th>
<th>Based on L+B+</th>
<th></th>
<th></th>
<th>Based on L+B-</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>For</td>
<td>Against</td>
<td>For</td>
<td>Against</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-propositional: 26* 11*

Note: Logical conclusion is marked by an asterisk.
frequency of conclusions for this problem type against the bias. Although there is slight variation over moods and particularly problem types, at this point, the bias is best viewed as a general tendency which may be constrained or augmented by validity (depending on the logical structure of any given problem). The relationship between the sub-superset directionality bias and other biases is taken up in the general discussion of this chapter.

GENERAL DISCUSSION

A strong effect of belief bias was found in Experiments 1 to 3, yet using identical problem content, the effect is not apparent on the syllogistic construction task. Differential response patterns on evaluation and construction tasks were, to some degree, implied by the analysis of concurrent verbalisations, collected in Experiment 3. These analyses suggested that conclusion-based (backward) processing was strongly associated with belief-based responding. It was further suggested that drawing attention to a specific conclusion of obvious truth status may encourage conclusion-based reasoning.

In the light of Experiments 1 to 3, the findings of Experiments 4 and 5 suggest that, with this type of problem content, the belief bias effect is paradigm specific. It occurs on the evaluation, but not on the construction task. However, since Experiments 4 and 5 were conducted, Oakhill & Johnson-Laird (1985) have, in fact, demonstrated a belief bias effect on the syllogistic construction task. This was more marked for indeterminate premise pairs, and was restricted to premises
yielding definitionally as opposed to empirically false conclusions. The distinction between empirical and definitional truth status echoes the findings of Revlin, Leirer, Yopp & Yopp (1980), who reported belief bias on the multiple choice task, but only with conclusions of definitional as opposed to empirical truth status.

Why should empirical belief bias be peculiar to the evaluation task? Comparing the evaluation task with the two alternatives studied we find important differences. The evaluation task requires the subject to consider one specific conclusion and no other. The presentation format focuses attention on one mood, one order of terms and one believable or unbelievable conclusion. The multiple choice and construction tasks do not focus attention to this extent. In the multiple choice task, although the ordering of terms is frequently unidirectional, subjects are presented with a variety of moods, and a non-propositional conclusion. Most importantly, this entails the presentation of a selection of conclusions whose truth status varies. The construction task is the least constrained of all. In terms of presentation, every permutable conclusion is as salient as any other. If syllogistic reasoning frequently proceeds in a backward manner, the search for alternatives is likely to be influenced by the potential conclusions which spring to mind. Since a conclusion of definitional truth status contains synonymous terms, whereas an empirically true or false one merely contains positively associated terms, it seems reasonable to class the former as more salient or available. If a conclusion is inherently available it
need not be dependent on task presentation to be salient. Conclusions having empirical truth status may not be sufficiently available in the construction task. They may need to be prompted by means of task presentation before they become a salient feature of the task. This explanation appears to be entirely consistent with Pollard's (1982) distinction between two types of availability effect - one of which is derived from the subject's experience, the other derived from salient features of task presentation. In the present instance, it is argued that some types of belief (i.e. empirical) are relatively less available than others (i.e. definitional). In such cases their salience needs to be increased by such factors as task presentation in order to obtain a belief bias effect. Although the availability explanation appears to fit the results so far, we should be wary of adopting it as a complete account of belief bias. Other results do not necessarily imply that belief bias is a function of the cue and not a process, as Pollard argues. Indeed, the findings of Revlin et al (1978, 1980) and Oakhill & Johnson-Laird (1985) suggest that the belief bias effect is not as straightforward as availability would predict. It seems that when a conclusion, shown to be dominant on an abstract version of the task, is unbelievable, subjects do not show a preference for a believable and hence more available alternative. Instead they prefer the non-propositional conclusion. It is difficult to envisage how this response could be cued by availability.

The proposal that certain types of effect may be paradigm specific is by no means a new one. Chapman & Chapman (1959) have suggested that the atmosphere effect may be induced by the
evaluation task. Indeed, the results of Experiments 4 and 5 tend
to support this contention. In both experiments, the mood of
conclusion predicted by atmosphere was not the most dominant
overall. Atmosphere predictions fared well on the figure 2
problems (AE, EA and EE premises). For both abstract and thematic
premises, atmosphere predicts the majority of correct responses
to determinate and the majority of errors on indeterminate
premise pairs. However, logic predicts more responses overall.
Turning to figure 3 problems for determinate premise pairs, the
dominant response is the (unbelievable) E conclusion. This is the
case for both thematic and abstract problems. Next come the
non-propositional and O conclusions. Clearly this does not
support the atmosphere prediction, of a predominance of O
conclusions. Atmosphere is supported slightly more on
indeterminate premise pairs, at least for abstract problems.
However, the logically correct, unpredicted, non-propositional
conclusion lies close behind the predicted O conclusion. For
indeterminate thematic problems the non-propositional conclusion
constitutes the dominant response.

For both Experiments 4 and 5 the response profile over the
five conclusion moods is more diverse than we might expect on the
basis of previous findings (see Johnson-Laird & Steedman, 1978).
Performance on abstract and thematic problems is similar,
however. In accounting for this variation between studies we must
bear in mind that the majority of variation is observed on the
figure 3 problems, which are notoriously difficult in this case.
Considering the small number of problems received by each
subject, it is possible that this may have prevented subjects
from developing the skill necessary to handle the relatively unwieldy amount of information presented.

Although there is no evidence of belief bias in Experiments 4 and 5, there is another semantically based effect. In both experiments there was a tendency to place subset before superset when producing a conclusion. There is no obvious explanation for this preference, although the significance of word order has been emphasized by several theorists. Johnson-Laird (1968) has drawn our attention to specific hypotheses which stress that word order is used to emphasize importance. Jespersen (1924), for instance, talks of "the centre of interest". Frege (1879) refers to "what we want the hearer to attend to specifically" and Clark (1965) claims that "people put what they want to talk about......in the beginning of a sentence". More recently, as we noted in Chapter 2, section 1, Begg & Harris (1982) have drawn attention to the principle of asymmetry. They claim that when a term is introduced as a subject or a predicate it conveys information beyond the syntactic functions of the terms. Specifically, hearers encode later information in terms of earlier focal information (Olson & Filby, 1971). Similarly, Hintikka (1973) suggests that such a statement presupposes specific fields of search. Thus, 'Some A are B', suggests a field of search throughout A, whereas the converse would imply that the search field would be B. As Begg & Harris point out, however, "in formal logic, the user is not entitled to assume that subject terms are more salient."

The directionality bias observed here has far reaching implications when placed in the context of other known biases. For instance, it may be that the sub-superset bias is directly
linked to the figural effects shown by Johnson-Laird & Steedman. If it is the case that listeners infer that predicate or subject terms are so placed because one represents 'focal' or 'given' information, then it follows that, since the conclusion is merely a restatement of the premises, this information will appear in the same position in the conclusion produced (if it is an end term). An extension of this argument, therefore, suggests that in the abstract syllogistic construction task, particular patterns will emerge. Thus, for figure 4 and figure 1 syllogisms, the A and C terms respectively will appear as the subjects of the conclusion. Since, for figure 2 syllogisms, both the A and the C terms appear as subjects in the premises and for figure 3 syllogisms neither term appears in this position, no directionality preference would be observed in conclusion production. Naturally this does not imply that directionality preferences would never occur on figure 2 and figure 3 syllogisms. Indeed, the directionality preferences observed in Experiments 4 and 5 showed just this. The crucial point here is that the content of the problems used was meaningful. In such cases, the syntactic position of terms is not the only determinant of this extra-logical inference. The semantic relationship between terms may also provide clues for the subject to act upon. The inclusion of specific terms such as 'police dogs' in contrast with a relatively general term like 'highly trained dogs' within the same premise pair may imply, as suggested earlier, that the specific term has been singled out for special treatment. The specific, or subset, term thus becomes the focal point of the statements made in the premises.
Consequently, the subject places this term before the superset term when producing the conclusion.

CONCLUSIONS AND IMPLICATIONS

Experiments 4 and 5 demonstrate that belief bias does not occur on the syllogistic construction task when the conclusions yielded are empirically true or false. This accords with the findings of Oakhill & Johnson-Laird who examined type of belief as a factor in syllogistic construction task experiments. Distinguishing between empirical and definitional truth status, they observed an effect of belief with the latter, but not the former problem content. Since Experiments 1 to 3 showed a strong and consistent empirical belief bias effect on the syllogistic evaluation task, it seems that certain types of belief bias may be induced by task characteristics. Whether or not a given belief will influence reasoning appears to depend on its salience. An empirical belief may require a certain type of task presentation which focuses attention on the conclusion to enhance its salience. The lack of any atmosphere effect on the construction task suggests that this, feature matching, tendency may be similarly induced.

The differential effect of empirical belief on the evaluation and construction tasks, along with the relatively pervasive effect of definitional belief, supports the view that reasoning is often guided by salience. Pollard's (1982) availability theory seems an appropriate framework in which to place the findings so far. However, this explanation is limited to research with empirical beliefs. It cannot account for the
pattern of responding shown with definitional beliefs on the multiple choice and construction tasks. There are alternative frameworks which can accommodate the results obtained. The mental models theory (Johnson-Laird, 1983) has been successfully applied to belief bias by Oakhill & Johnson-Laird. Evans' (1984) heuristic and analytic processing theory emphasizes the role played by the salience of task features and although it has not been directly applied to belief bias in syllogistic reasoning it appears to be particularly apt in the light of its ability to account for related content and presentation effects.

The hitherto undetected directionality bias in conclusion production, characterised as a tendency to place sub before superset was linked to Begg & Harris's (1982) principle of asymmetry. This may be linked to the figural effect, which could be seen as a tendency to place terms in the position (subject or predicate) which best retains the assumed extra logical implications of the premises. This does not necessarily question existing theoretical accounts of the figural effect. The analogical (Johnson-Laird & Steedman, 1978) and mental models (Johnson-Laird, 1983) explanations claim that figural effects result from the structure of representational formats, and (in the latter case) a first in, first out principle of working memory. Both could be influenced by, or reflect, linguistic presuppositions with this type of task. Viewed in this way, the figural effect can be associated with other linguistic directionality effects considered in Chapter 2 section 1. Both linguistic and semantic features of the task appear to invite extra logical inferences, directing reasoning. Hence, it may be
theoretically profitable to view the resultant effects as related phenomena.
Chapter 7
The logic x belief interaction - the search for an explanation.

General Introduction 194

Experiment 6
Introduction 198
Method & Design 199
Results & Discussion 204
Calculation of logic, belief and interaction indices. 205

Experiment 7
Introduction 210
Method & Design 211
Results & Discussion 212

Experiment 8
Introduction 216
Method & Design 216
Results & Discussion 217

Experiment 9
Introduction 220
Method & Design 223
Results & Discussion 227

General Discussion & Conclusions 229
INTRODUCTION

So far we have established that, with a particular set of materials, belief bias is strong and consistent on the syllogistic evaluation, but not the syllogistic construction task. Moreover, along with this effect there is a strong effect of logic and a logic x belief interaction.

There are at least three different ways to explain this interaction. The simplest explanation is that the strength of conflict arising from believable and unbelievable conclusions was not equivalent with the materials used in Experiments 1 to 3. On average, unbelievable conclusions were rated only one point away from the mid-point of the believability scale. In other words, such statements were rated as only slightly unbelievable. Believable conclusions were, on average, rated at the extreme end of the scale and hence as highly believable. Thus, it is possible that the unbelievable conclusion, being weaker in terms of belief, exerts a weaker effect. This may result in subjects regarding such conclusions as neutral. Without being able to use belief as an indicator people may resort to checking the logical validity of a conclusion.

An alternative, intuitively appealing, explanation of the logic x belief interaction is that subjects have a tendency to accept, uncritically, conclusions with which they agree. They are more likely to check the validity of conclusions which conflict with their beliefs. A related study by Lord, Ross & Lepper (1979) tends to support this interpretation. Their study investigated the effects of prior theories on the evaluation of new evidence. Students were asked to indicate their belief about the deterrent
effects of capital punishment. They were then required to consider the merits of two purportedly authentic studies, on the threat of capital punishment. In a counterbalanced design, each subject read both a study which supported their own position and a study which opposed it. The former was reported by subjects to be more convincing and better conducted than the other study they had read about. Where there was agreement between study and subject, flaws were allowed to pass apparently unnoticed. Where there was conflict, severe criticism followed. Apparently, the treatment of evidence addressing an existing belief is sensitive to the conclusion reported. Evidence appears to be selectively scrutinised, according to the believability of the conclusion.

The other way in which to characterise the logic x belief interaction is to say that belief bias appears stronger on invalid problems. We noted earlier that the premises of the invalid problems used in Experiments 1 to 3 never actually contradict the conclusion given. In other words there is no inherent falsification in such cases. Logically, the conclusion is invalid because it does not necessarily follow. However, there may be a difference here between the logical conception of necessity and that used by subjects. Recent studies by Dickstein (1980,1981) have suggested that conclusions are often deemed to be valid so long as they are not false, that is, so long as they are possible. In Experiments 1 to 3 some subjects may simply be evaluating invalid conclusions according to belief because they are not demonstrably true or false on the basis of the premises alone. If such conclusions are possible and believable, they are accepted. If they are possible, but seem intuitively false, then
they are rejected. According to this explanation, subjects consider both validity and belief, but respond differently to different types of conflict. On this basis we would expect instructions emphasising logical necessity to reduce the interaction. Indeed, when such instructions were used, as in Experiment 3, the interaction failed to reach significance.

A fundamental difference between the latter explanations lies in the level at which the 'logical' analysis takes place. The penultimate (selective scrutiny) explanation is based on an essentially non-logical believability assessment, followed in some cases by a more questioning premise analysis. The misinterpreted necessity explanation is based on an essentially logical approach which incorporates an illogical component. Subjects misinterpret the logical meaning of necessity. It is this misunderstanding which leads to a belief based decision in some instances.

One way of achieving an empirical distinction between the two explanations is to give subjects invalid problems which have a demonstrably false conclusion, i.e. a conclusion which is directly contradicted by the logical premises. Under such conditions the selective scrutiny model would predict no reduction in the logic x belief interaction. According to this model the interaction is due to the unquestioning acceptance of believable conclusions and critical evaluation of unbelievable conclusions. Since the conclusion evaluation in this model is not subject to distortion, due to misunderstanding the concept of logical necessity, the aid to analysis provided by the demonstrably false conclusion would not alter the overall pattern
of responding. The misinterpreted necessity model, on the other hand, predicts that this would serve to obviate the misunderstanding and hence remove the interaction.

Unfortunately, there are several problems with this method of testing. In order to make strict comparisons between test problems and those used in Experiments 1 to 3, the structural identity of the originals should be retained along with any controls built in. The original premises were convertible and the conclusions given accorded with the atmosphere produced. No such problems exist which have a demonstrably false conclusion.

An alternative method of testing the misinterpreted necessity model is to give subjects an explanation of logical necessity in the instructions. Experiments 6 to 8 examine this approach.
EXPERIMENT 6

Experiment 6 had three main aims. The first was to test the asymmetrical believability explanation of the logic x belief interaction. This is the simplest of the three accounts put forward and claims that the interaction is due to unbelievable conclusions being only slightly unbelievable and believable conclusions being completely believable, on average. This results in belief bias being stronger on believable than unbelievable problems. This explanation was tested by changing the problem contents to those which had more extreme believability ratings for unbelievable conclusions.

The second aim was to test the logic based, misinterpreted necessity model of the logic x belief interaction. This involved manipulating instructions. Instructions were of two types: Standard and Augmented. The latter included an explanation of how to interpret necessity; the former did not. Assuming that subjects understand and comply with their instructions, the misinterpreted necessity model predicts no logic x belief interaction under Augmented instructions. This is because the basis of the interaction - the misinterpretation - has now been cleared up. The selective scrutiny model would still predict the interaction, regardless of instruction type. Evidence suggests that subjects can understand and comply with instructions which explain necessity (see Dickstein, 1981), and this seems to reduce incorrect responses to some indeterminate problems. The manipulation of instructions, therefore, appeared to be a reasonable test of the misinterpreted necessity model.
The third aim was to establish the nature of belief bias. Several theorists have tended to assume that belief bias is simply a tendency to reject unbelievable and accept believable conclusions (e.g. Janis & Frick, 1943). There is no unequivocal empirical evidence on which to base this assumption and since the misinterpreted necessity model and selective scrutiny models posit conflicting characterisations of belief bias it seems that this question should be clarified. The misinterpreted necessity model claims that only certain types of (indeterminate or inconclusive) problems receive a belief based judgement. Alternatively, the selective scrutiny model claims that belief selectively influences the analysis of believable and unbelievable problems in different ways. Specifically, with this task, there is a pure belief-based evaluation of believable conclusions. For unbelievable conclusions the belief in a conflicting conclusion motivates the subject to scrutinize the premises. One way to discover whether there is, in fact, a simple tendency to both reject and accept on the basis of belief is to compare performance on controversial with that on neutral problems. Although such attempts have been made in the past (e.g. Lefford, 194C; Kaufman & Goldstein, 1967; Nehrke, 1972), the results obtained are difficult to interpret due to methodological flaws (see Chapter 4, section 1).

METHOD

Materials

The mood and structural composition of problems were identical to those used in Experiment 2. Subjects were given a
new collection of problem contents which were selected to provide highly believable, highly unbelievable and neutral conclusions. This enabled a test of the asymmetrical believability explanation of the logic x belief interaction and provided a neutral problem set with which to compare performance on controversial problems. The believability of all conclusions was ascertained by questionnaire. As in earlier experiments, subjects who took part in this exercise were excluded from participation in the experiment proper. The rating scale differed from that used in previous questionnaires. In this case the 7-point scale ranged from -3, through zero to +3 (where -3 was completely unbelievable, zero was neutral and +3 was completely believable). This alteration was made to create a clearer distinction between believable, neutral and unbelievable in the subject's mind. The believability ratings for the four believable, four neutral and four unbelievable statements used as conclusions in the experiment are given in Appendix IV.

Great difficulty was encountered in obtaining neutral (i.e. zero) ratings for statements which the experimenter, at least, deemed to be neutral. Although the statements chosen for use in the experiment stray, on average, only marginally from the zero mid-point several of the other statements included as neutral in the questionnaire were given a rating of 'slightly believable'. This finding casts doubt on the 'neutral' statements used in other experiments which did not employ an objective measure of neutrality.
Design

Subjects were divided into two groups: Group S (which received 'Standard' instructions) and Group A (which received 'Augmented' instructions, containing an explanation of logical necessity). 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.

Combination of problem type and problem content was such that each content occurred an equal number of times with each problem type. Presentation order was randomised. Problems were presented in booklet form. Responses were written in a space provided below each conclusion. Each problem appeared on a separate page.

Subjects

Forty-eight undergraduates at Plymouth Polytechnic took part in partial fulfilment of a course credit requirement.

Procedure

Subjects were run in groups of four. Each subject was given unlimited time to evaluate the validity of six syllogisms.
Instructions

Standard instructions were based on those given in Experiment 3 (with references to the prose passage and verbalisation removed). Augmented instructions simply contained an additional passage outlining the principal of logical necessity with a short reminder at the end. These instructions were designed to follow a specific section of the instructions given by Dickstein (1981). (Dickstein's instructions were not included in his publication. They are given in Appendix V).

The instructions were as follows:

Standard Instructions

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 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 in front of you in case you need to refer to them later on.

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.

**Augmented Instructions**

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 in 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 YOU SHOULD 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 7.1 sets out the response patterns found for the Standard and Augmented instruction groups.

Sign tests on controversial problems yielded the following results. For both instruction groups there was a substantial effect of logic ($p < 0.001$ in both cases). A substantial effect of belief was also observed, but for the Standard instruction group only ($p = 0.011$) and an interaction between logic and belief was also found for this group ($p = 0.011$). All tests were one-tailed.
Sign tests on neutral problems showed a highly significant effect of logic for both groups ($p < 0.001$ in both cases, one-tailed tests). Overall, accuracy on neutral problems was 79%. Accuracy on invalid problems was only 13% lower than on valid problems. Table 7.1 clearly shows no apparent difference between the Standard and Augmented instruction groups with neutral problems.

In order to compare performance between groups, three Mann-Whitney tests were performed on logic, belief and interaction indices, calculated for each subject. Indices were calculated in the following manner:-

<table>
<thead>
<tr>
<th>INDEX</th>
<th>CALCULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logic</td>
<td>Total number of acceptances for invalid problems subtracted from that for valid problems.</td>
</tr>
<tr>
<td>Belief</td>
<td>Total number of acceptances for unbelievable from that for believable problems.</td>
</tr>
<tr>
<td>Interaction</td>
<td>Total number of acceptances for the Valid-Believable problem, plus that for the Invalid-Unbelievable problem. This is then subtracted from the total number of acceptances for the Valid-Unbelievable plus Invalid-Believable problem.</td>
</tr>
</tbody>
</table>

Comparisons between Standard and Augmented instruction groups on each of the three (logic, belief and interaction) indices revealed only one significant difference, which was for belief indices ($p < 0.05$, one-tailed test).
Table 7.1

The percentage frequencies of subjects accepting the conclusion, i.e. deeming it to be valid, are shown for each problem type, divided according to instruction group. Experiment 6 (n=24 in each instruction group).

<table>
<thead>
<tr>
<th></th>
<th>Believable</th>
<th>Unbelievable</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>75</td>
<td>75</td>
<td>83</td>
</tr>
<tr>
<td>Invalid</td>
<td>50</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td><strong>Augmented</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>71</td>
<td>79</td>
<td>88</td>
</tr>
<tr>
<td>Invalid</td>
<td>17</td>
<td>4</td>
<td>25</td>
</tr>
</tbody>
</table>
Let us firstly consider the results with regard to the asymmetrical believability explanation. Recall that the problem content was changed from that employed in Experiments 1 to 3. In this experiment, unbelievable conclusions were rated as completely unbelievable. Under such circumstances the rating of believable and unbelievable conclusions is largely symmetrical. Hence, the asymmetrical believability explanation would not predict an interaction for the Standard instruction group. Since a logic x belief interaction is clearly evident for this group we must reject this explanation.

The misinterpreted necessity model receives some, albeit limited, support. Sign tests indicate that the logic x belief interaction is not significant under Augmented instructions, as predicted. However, the Mann-Whitney test comparing interaction indices for the two instruction groups shows no significant difference between them. Interestingly, there was no belief bias effect for the Augmented instruction group. Examination of Table 7.1 suggests that the major difference between the Standard and Augmented instruction groups lies in responses to invalid problems. Neither group appears to show any effect of belief on valid problems. If anything there is a reverse tendency under Augmented instructions.

A further problem for the misinterpreted necessity model concerns the comparative behaviour of neutral problems under Standard and Augmented instructions. According to this model, performance should improve on invalid problems under Augmented instructions. There is, in fact, a very slight, but non-significant improvement for both valid and invalid problems.
This failure to improve could be due to a ceiling effect on Standard instructions, but this raises the question of why the ceiling effect should occur in the first place. This type of problem should be one of the most difficult to solve.

Thus far we have rejected the asymmetrical believability explanation and found limited support for the misinterpreted necessity model. We now turn to the third aim: to discover the nature of belief bias. Sign tests on valid and invalid problems, comparing performance on controversial with neutral conclusions, revealed no significant differences between believable and neutral problems. This was the case for both instruction groups with both valid and invalid problems (p > 0.05 in all cases, one-tailed tests). However, a significant difference was observed between unbelievable and neutral problems, but for invalid problems only. This difference was significant for both instruction groups (Standard p < 0.01, Augmented p < 0.05. One tailed tests in each case). Since the ratings for 'neutral' conclusions were, on average, narrowly believable it may be unwise to place too much emphasis on the outcome of this particular statistical analysis, as there may be a bias towards 'neutrals' being treated as believable, even if only very slight. An examination of the overall pattern of responding may be more informative. Focusing on the Standard instruction group, we find that the majority of variation in acceptance rates occurs on invalid problems. For unbelievable problems there is 100% correct performance, for neutrals it is 71%, and on believables it is 50%. Hence, compared with neutrals, unbelievable problems lead to fewer errors and believable problems led to more. There could be
two reasons for this. Firstly, as the misinterpreted necessity model suggests, people reject and accept invalid problems on the basis of the believability of the conclusion. Secondly, as the selective scrutiny model suggests, people accept believable problems uncritically, but subject unbelievable problems to critical analysis. Neither model predicts a belief bias on valid problems - which, in fact, there is not. On the basis of these results it appears that there is no overall tendency to accept and reject on the basis of belief - there seems to be no evidence of 'rejection of the logical task' in this experiment.

We must conclude that the misinterpreted necessity model has not received unqualified support from this experiment. The failure to obtain adequate supportive evidence could be due to numerous factors. Various features of the experimental results are puzzling. The loss of belief bias on valids, the reverse belief bias tendency observed on valid problems with Augmented instructions and the lack of improvement on neutral problems with Augmented instructions seem inexplicable at this point. The use of new problem content and the intermingling of neutral with controversial content, within subjects, could have distorted the results. Such possibilities are examined in Experiment 7.
EXPERIMENT 7

This experiment had three main aims. The first was to replicate the findings of Experiment 6, using larger subject numbers. The failure to obtain conclusive evidence in support of the misinterpreted necessity model in Experiment 6 could have been due to the relatively small number of subjects used. If this is the case we would expect to find a more marked effect of instructions in Experiment 7.

The second aim was to assess the generalisability of effects across problem contents, comparing that used in Experiment 6 with a selection of that used in earlier experiments. In Experiment 6, the distinction between believable and unbelievable conclusions was more marked than that for all previous experiments in this series. We cannot assume that belief bias is equivalent across all types of problem content. In Experiment 7 all subjects received both types of problem content. These are referred to as original and new contents.

The third aim concerned the possible effects of intermingling controversial with neutral problem content. Lefford's (1946) study suggested that carry-over effects may occur, but only when each type of content is presented in blocks. In Experiment 6 neutral problems were randomly intermingled with controversial problems. On the basis of Lefford's findings no distortion of response patterns for either problem type would be expected. However, there are paradigm as well as methodological differences between the two studies. In order to discover whether the results of Experiment 6 had been distorted by this design feature, subjects in Experiment 7 were
METHOD

Materials

The mood and structural composition of all problem types were identical to those of Experiment 6. Two sets of problem contents were used. One set consisted of original thematic content, the other consisted of new thematic content (as used in Experiment 6).

Design

Subjects were divided into two groups: Group S (which received Standard instructions) and Group A (which received Augmented instructions). Each subject was asked to evaluate eight syllogisms, four of original and four of new content. Problems were presented in a booklet which contained two blocks of four problems. Half of each group received booklets with a block of original followed by a block of new problem content. The remaining half of the group received the reverse ordering of blocks. For each block of problem content subjects received each of the four problem types. Valid (Believable, Unbelievable) and Invalid (Believable, Unbelievable). Combination of problem type and content was identical to that of previous experiments. Presentation order was randomised within blocks.
Subjects

Sixty-four undergraduates at Plymouth Polytechnic took part in partial fulfilment of a course credit requirement.

Procedure

Subjects were run in groups of four. Each subject was given unlimited time to evaluate the validity of eight syllogisms.

Instructions

As for Experiment 6

RESULTS AND DISCUSSION

Table 7.2 sets out the percentage frequency of conclusions accepted, divided according to instructions and problem content. Response patterns across contents appear to be similar.

The first task was to discover any differences due to problem content, original or new. Comparing both types of content for each instruction group, no significant differences were found for logic, belief or interaction indices. (Mann-Whitney tests, two-tailed.) It therefore seems that original and new problem contents did not have differential effects on overall acceptance rates for the four problem types. However, a within-subject design was used and overall response patterns could have been distorted by carry-over effects. Since all subjects received a block of both problem contents, the two content blocks in each booklet were separated. Examination of response patterns for the
first block of content to be presented showed close similarities to those shown in Table 7.2. Hence, the overall findings do not appear to have been distorted by any carry-over effects. Consequently, all further analyses were carried out on data for both original and new contents combined.

Sign tests revealed significant effects of logic, belief and a logic x belief interaction for both instruction groups (p < 0.001 in all cases, except for the Standard instruction group interaction, where p < 0.01. All one-tailed tests). Mann-Whitney tests on logic, belief and interaction indices comparing Standard and Augmented instruction groups revealed no significant differences (p > 0.05 in all cases, one-tailed tests). Table 7.3 sets out the overall acceptance rates to provide a clear picture of the patterns found for each instruction group.

The results of this experiment are clear cut. Original and new problem contents do not appear to have differential effects on performance. This finding is not unexpected. The two types of content differed on one dimension only; the type of unbelievable conclusion presented. Both types of content were otherwise very similar. Both were chiefly concerned with definition of truth status as opposed to empirical truth status.

The most striking finding, in terms of the model being tested, was the absence of any significant differences between instruction groups. Crucially, both instruction groups showed a strong logic x belief interaction. This is clearly contrary to the prediction made by the misinterpreted necessity model. Hence, the only evidence in support of this model comes from Experiment...
Table 7.2
Percentage frequency of conclusions accepted for the four problem types, divided according to content (Original or New) and Instructions (Standard or Augmented). Experiment 7 (n=32 in each instruction group).

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th></th>
<th></th>
<th>Augmented</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>Original content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>84</td>
<td>78</td>
<td></td>
<td>78</td>
<td>84</td>
</tr>
<tr>
<td>Invalid</td>
<td>63</td>
<td>13</td>
<td></td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>New content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>84</td>
<td>81</td>
<td></td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>Invalid</td>
<td>47</td>
<td>19</td>
<td></td>
<td>53</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 7.3
Percentage frequency of conclusions accepted for the four problem types, divided according to instruction group (original and new contents combined). Experiment 7 (n=32 in each group).

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th></th>
<th></th>
<th>Augmented</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>Valid</td>
<td>84</td>
<td>80</td>
<td></td>
<td>81</td>
<td>84</td>
</tr>
<tr>
<td>Invalid</td>
<td>55</td>
<td>16</td>
<td></td>
<td>44</td>
<td>5</td>
</tr>
</tbody>
</table>
6, and this is somewhat limited. Response patterns shown in Table 7.3 do seem to differ slightly from those shown in Table 7.1, for Experiment 6. With regard to invalid problems under Augmented instructions, the difference between believable and unbelievable problems is less marked in responses for Experiment 6. Before placing too much emphasis on this, the reliability of the pattern found in Experiment 6 (Augmented instructions) should be established by a replication.
EXPERIMENT 8

Experiment 7 found a strong logic x belief interaction under Augmented instructions and this contradicts the claim made by the misinterpreted necessity model. Experiment 6 stands alone in suggesting a trend in the predicted direction. There are two explanations for these conflicting results, although both have ominous implications for the survival of the misinterpreted necessity model. Firstly, the neutral problems included in Experiment 6 could have altered response patterns - resulting in a fall in the logic x belief interaction when combined with Augmented instructions. This would imply that the interaction is not solely due to the misinterpretation of necessity. Alternatively, the loss of a logic x belief interaction under Augmented instruction, in Experiment 6, could have been due to a Type II error.

Experiment 8 was designed to focus on the Augmented instruction group of Experiment 6. Its purpose was to establish whether or not this aspect of experimental findings was replicable. Hence, subjects received both controversial and neutral problems.

METHOD AND PROCEDURE

The materials and syllogisms used were identical to those of Experiment 6. The experimental design differed from Experiment 6 in that no Standard instruction group was included.

Subjects

Seventy-two undergraduates at Plymouth Polytechnic took part in partial fulfilment of a course credit requirement. They were tested in groups of four.
RESULTS AND DISCUSSION

The pattern of responses over the six different problems is set out in Table 7.4.

Sign tests revealed a highly significant effect of logic ($p < 0.001$). This replicates the strong effect of logic found in Experiment 6. In contrast to Experiment 6, a significant effect of belief was found ($p < 0.001$). Moreover, the logic x belief interaction was marginally significant this time ($p = 0.08$) All one-tailed tests.

The pattern of results for the Augmented instruction group of Experiment 6 was not replicated in this experiment. The results failed to substantiate the claims of the misinterpreted necessity model. The failure to replicate, the nature of results, together with the results of Experiment 7 provide a strong basis on which to reject this model of the logic x belief interaction.

The findings of this experiment, combined with those of Experiment 6 appear to question the very foundations of the misinterpreted necessity model. Dickstein (1981) claims that when necessity is explained, performance on this type of syllogism improves. Experiment 6 showed no such improvement on neutral syllogisms when responses under Standard and Augmented instructions were compared (see Table 7.1). However, in both cases, the frequency of correct responses to both valid and invalid syllogisms was high. Hence, the lack of improvement on invalids under Augmented instructions could have been due to a ceiling effect. However, in Experiment 8, performance on neutrals
Table 7.4
The percentage frequency of subjects accepting the conclusion, i.e. deeming it to be valid. Experiment 8 (n=72).

<table>
<thead>
<tr>
<th></th>
<th>Believable</th>
<th>Unbelievable</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>72</td>
<td>64</td>
<td>75</td>
</tr>
<tr>
<td>Invalid</td>
<td>35</td>
<td>8</td>
<td>43</td>
</tr>
</tbody>
</table>
is not so impressive. Despite the Augmented instructions, the percentage correct is only 57%, which is only slightly better than chance level. Although evidence from Experiments 6 and 8 is limited, the findings suggest that instruction on the interpretation of necessity does not produce the improvement in performance levels indicated by Dickstein. Apparently Dickstein's results do not generalise across paradigms and or problem contents.
EXPERIMENT 9

Augmented instructions did not have the effect predicted by the misinterpreted necessity model. However, they did in fact, alter response profiles. Table 7.5 sets out the response profiles for those experiments which included both Standard and Augmented instructions.

There is clearly a fall in acceptance rates for Invalid-Believable problems under Augmented instructions, but there is also a reversed belief bias on valid problems. The reason for this is not entirely obvious. A re-examination of the Augmented instructions suggests an intriguing possibility. The paragraph included to explain logical necessity could also be viewed as a subtle warning against accepting conclusions merely because they are believable. The fall in acceptance rates for Valid-Believable problems is less than that for Invalid-Believable problems so this cannot simply be a tendency to reject believable conclusions. Rather, it seems that believable problems are being subjected to greater logical scrutiny than is the case under Standard instructions. If some subjects did, in fact, interpret the Augmented instructions as a subtle hint not to accept conclusions just because they are believable, the misinterpreted necessity model could still not account for the reverse belief bias on valids and would still predict a loss of the logic x belief interaction (the Invalid-Believable cell would be the only one affected). The relevant portion of the Augmented instructions is repeated below to allow the reader to judge the feasibility of the suggested
Table 7.5

Percentage frequency of conclusions accepted in Experiments 6 and 7, divided according to type of instructions.

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Augmented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>Experiment 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Invalid</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Experiment 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>84</td>
<td>80</td>
</tr>
<tr>
<td>Invalid</td>
<td>55</td>
<td>16</td>
</tr>
</tbody>
</table>
interpretation.

"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'."

If Augmented instructions led to less reliance on belief and greater logical scrutiny on believables, this implies that instructional factors may play a part in controlling what is subjected to analysis when reasoning. In the series of experiments reported here, the instructions which deviated most from Standard instructions were the very complex instructions given to accompany the construction tasks of Experiments 4 and 5. Notably, these instructions strip the syllogism to its bare bones. They emphasize that it is structure and not content which determines the conclusion. These instructions give an example syllogism which is abstract in content and they explain that the conclusion is derived from the structural relationships expressed in the premises. Hence, such instructions could be drawing the subject's attention to logical structure rather than content (believability specifically). It follows from this that one factor contributing to the loss of belief bias on the construction task may have been the instructions drawing attention to logical structure.
Experiment 9 was designed to discover what effect these instructions might have when coupled with an evaluation task. This simply entailed using two groups of subjects, one receiving "Simple" instructions, the other receiving "Complex" instructions, based on those given in Experiments 4 and 5. Simple instructions were simply a shortened version of Complex instructions with the emphasis on structure reduced and the example syllogism removed. Performance on the four problem types was then compared.

METHOD

Design

All subjects received four syllogisms to evaluate, which consisted of the four problem types: Valid-Believable, Valid-Unbelievable, Invalid-Believable and Invalid-Unbelievable. All problems were in figure 3 format. The problem contents were identical to those used in Experiment 1. Each problem type occurred equally with each problem content and presentation order was randomised. Problems were presented in booklet form and a space was provided beneath each conclusion for responses (Yes or No) to be written.

Subjects

32 undergraduates at Plymouth Polytechnic acted as paid volunteers. Subjects had no previous experience of this task, nor any training in logic.
Procedure

Subjects were run in groups of four and divided into two groups. One group received Simple instructions. The other group received Complex instructions. The complete instructions were as follows:

Simple Instructions

This experiment is designed to find out how people solve logical problems. You will be tested on 4 logical reasoning problems, which are contained within the booklet which you have been given. Your task is to decide whether or not a given conclusion follows logically from the information given - and this information only. You must assume that all the statements within the problem are true - this is very important. If, and only if, you judge that the given conclusion logically follows from the statements given you should answer by writing "YES" below the conclusion, otherwise write "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 after you have started.

Please keep these instructions in front of you in case you need to refer to them later on.

REMEMBER, IF AND ONLY IF YOU JUDGE THAT THE GIVEN CONCLUSION LOGICALLY FOLLOWS FROM THE STATEMENTS GIVEN YOU SHOULD 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 aid you in this task.

**Complex Instructions**

This experiment is designed to find out how people solve logical problems. In the booklet which you have been given there are 4 logical reasoning problems. Your task is to decide whether or not the conclusion given does or does not logically follow from the information which is given above. The information takes the form of two statements (premises) which can be expressed symbolically as follows:

ALL B ARE A,
SOME C ARE B.

As you can see, the two premises tell us something about the relationship between three terms: A, B and C. The term B never appears in the conclusion; since the conclusion is a statement about the relationship between A and C, or vice-versa. The conclusion to the above example is, therefore, SOME C ARE A.

Since this a problem requiring logical analysis, you should interpret the word 'SOME' in its strictly logical sense; meaning **AT LEAST ONE AND POSSIBLY ALL**. So the statement "SOME B ARE C" does not necessarily also mean that SOME B ARE NOT C.

In the booklet you will find 4 different logical problems. They are the same type of problem as the example problem which is
shown above, however, the terms used will not be letters of the alphabet, but real words instead. Your task is to write down, below the conclusion given, 'YES' if you judge that the conclusion necessarily follows from the information given, or 'NO' if you judge that the conclusion does not necessarily follow from the information given.

You are reminded that you must base your decision on the information given in the two premises - and this information only. 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 specific conclusion logically follows from the information given you should write 'YES'; the conclusion given may not always be the correct one.

Please take your time and be certain that you have made the logically correct decision 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 in front of you in case you need to refer to them later on.

REMEMBER, YOUR DECISION SHOULD BE BASED SOLELY UPON WHAT CAN BE DEDUCED WITH ABSOLUTE CERTAINTY FROM THE TWO PREMISES - AND THIS INFORMATION ONLY

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 aid you in this task.
RESULTS AND DISCUSSION

Sign tests revealed an effect of logic under both Simple and Complex instructions (p < 0.05, one-tailed tests for both groups). An effect of belief was observed under Simple instructions (p < 0.05), although for Complex instructions the effect fell just short of significance (p = 0.055, one-tailed test). An interaction between the two effects was present under Simple instructions (p < 0.01), but no interaction was observed under Complex instructions (p > 0.05 one-tailed test). Table 7.6 presents the response profiles for the two instruction groups.

Comparisons between instruction groups revealed no significant differences for logic indices (p > 0.05). This was also true for belief indices (p > 0.05). However, there was a significant difference for interaction indices (p < 0.05). (All one-tailed Mann-Whitney tests.)

Although there were no significant differences on belief indices, belief bias fell short of significance under Complex instructions. Inspection of Table 7.6 shows a marked drop in belief bias here. This tends to support the contention that the Complex instructions contributed towards the loss of belief bias on the syllogistic construction tasks. With the evaluation task belief bias does not totally disappear. However, we must bear in mind that a belief judgement is being prompted by the explicit presentation of a conclusion of obvious truth status. In this case the subject's attention is being drawn to both structure (by the instructions) and truth status (by the nature of the task). In the construction task experiments there is no given conclusion. This reduces the likelihood that belief would be a
Table 7.6

Percentage frequencies of subjects accepting the conclusion for each of the four problem types, divided according to instruction group. Experiment 9 (n=16 in each group).

<table>
<thead>
<tr>
<th></th>
<th>Simple instructions</th>
<th>Complex instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Believable</td>
<td>Unbelievable</td>
</tr>
<tr>
<td>Valid</td>
<td>81</td>
<td>69</td>
</tr>
<tr>
<td>Invalid</td>
<td>81</td>
<td>19</td>
</tr>
</tbody>
</table>
In conclusion, it seems that the Complex instructions given in the construction tasks could have contributed towards the loss of belief bias. The effect may have been exacerbated by the nature of the task itself. In both cases attention is directed to logical structure rather than believability. The question of whether belief bias would occur on the construction task without instructions of this type would be dependent on the belief in question and its relative level of salience.

GENERAL DISCUSSION AND CONCLUSIONS

The experiments in this chapter were predominantly concerned with ascertaining the basis of the logic x belief interaction. In addition, Experiment 6 attempted to determine the nature of belief bias. This experiment could not provide a definitive answer to this question because the ratings for neutral conclusions strayed narrowly from the neutral to the slightly believable rating point. This rendered the interpretation of statistical analyses somewhat dubious. However, inspection of general trends for the Standard instruction group suggested that performance on valid problems was very similar for believable, unbelievable and neutral problems. For invalid problems there was considerable variation. Performance on unbelievable problems was best and it was worst for believable problems. The number correct for neutrals lay at a mid-point almost equidistant from the two. This suggests that both belief and disbelief influence performance. This conclusion is consistent with that of Janis & Frick (1943) and Kaufman & Goldstein (1967). However, other findings do not replicate those of Kaufman & Goldstein, in
particular, in Experiment 6 correct performance on neutral problems fell at an intermediate level. Kaufman & Goldstein found neutrals to be associated with the highest level of errors overall. However we should note that Kaufman & Goldstein did not have their neutral conclusions rated for believability. Given the problems encountered in trying to find non-affective statements for Experiment 6, we should be wary of accepting Kaufman & Goldstein's subjective categorization of neutrals. This might explain discrepancies concerning this factor. Kaufman & Goldstein report an order effect on valids as well as invalids. The frequency of errors was greatest for unbelievable, less for "neutral" and least of all for believables. No such pattern was observed in Experiment 6. In fact, there was no belief bias on valid problems.

Both Janis & Frick and Kaufman & Goldstein report a logic x belief interaction, though neither study explains why it should occur.

Of the three models of the logic x belief interaction considered in Experiments 6 to 8, the selective scrutiny model provides the most convincing account of results so far. The two models which were subjected to direct testing were not supported by the results obtained. The asymmetrical believability explanation was directly contradicted by results. The misinterpreted necessity model was shown to be lacking on two counts. The descriptive adequacy of the model was unsatisfactory with regard to Augmented instructions. Moreover, the lack of improvement on neutral problems under Augmented instructions seriously questions its explanatory adequacy. It seems that
Dickstein's (1981) explanation of errors on such invalid problems, does not, at the very least, generalise across paradigms and or content. The fault does not appear to lie with the Augmented instructions given here, as they were closely based on Dickstein's instructional explanation of necessity.

Overall, it seems that the most fitting explanation of the logic x belief interaction is that provided by the selective scrutiny model. Although a component of the belief bias effect may be attributable to rejection of the logical task as Henle (1962) claims, it seems that belief does not have an entirely detrimental effect on reasoning as others (e.g. Morgan & Morton, 1944) suggest. The selective scrutiny model claims that the interaction, and hence a component of the belief bias effect, is due to selective inhibition and facilitation of analysis on believable and unbelievable problems respectively. This model has implications which are not purely restricted to studies of belief bias. In Part 3 we will attempt to apply the selective scrutiny model directly to the results obtained in Experiments 1-9. We will then consider its generalisability and theoretical implications with particular reference to issues raised in the review.
SECTION THREE

GENERAL DISCUSSION

CHAPTER EIGHT
A re-examination of the belief bias effect in the light of Experiments 1 to 9 - steps towards a theoretical explanation.

233

CHAPTER NINE
The selective scrutiny model of belief bias.
An integration of empirical evidence and a consideration of other issues.

259
Chapter 8

A Re-examination of the belief bias effect in the light of Experiments 1 to 9 - steps towards a theoretical explanation.

8.1 Is there a belief bias? 234
8.2 What is the nature of belief bias? 239
8.3 Comparisons between neutral problems and those with obvious truth status. 240
8.4 The interpretation and implications of protocol data. 242
8.5 The logic x belief interaction (including the Misinterpreted Necessity model). 245
8.1 Is there a belief bias?

In order to gain a theoretical understanding of the nature and working of belief bias it is profitable to consider Experiments 1 to 9 with respect to questions raised in Chapter 4. A fundamental question concerns whether or not belief bias actually exists. If the inferences that people make are influenced by their beliefs, is the resultant effect sizeable enough to warrant concern?

The nine experiments reported here have focused largely on the syllogistic evaluation task. Problem contents have been restricted to those concerning empirical rather than definitional truth status. Experiments 1 to 3 demonstrated a strong and consistent effect of belief bias. An effect of logic was also observed along with a logic x belief interaction. The effect of belief was shown to be relatively strong compared with that of logic. For all three experiments combined, the percentage frequency of correct responses was 90% when logic and belief agreed. This fell to 43% when they conflicted. This pattern of results occurred in all three experiments. The response profile observed suggests that where errors occurred, they were largely in the direction of belief and that the logically correct response rate was elevated when belief was coincidental. This is shown in table 8.1.1 which presents the means for all three experiments combined.
Table 8.1.1

Percentage frequency of conclusions accepted for each problem type, for Experiments 1 to 3 combined.

<table>
<thead>
<tr>
<th></th>
<th>Believable</th>
<th>Unbelievable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>89</td>
<td>56</td>
</tr>
<tr>
<td>Invalid</td>
<td>70</td>
<td>9</td>
</tr>
</tbody>
</table>

This finding substantiates those reported in the majority of syllogistic evaluation task studies. With the exception of Frase (1966a,b;1968), most findings reveal a characteristic pattern of conclusion evaluation; whereby conclusions appear to be accepted or rejected according to their a priori believability. The problem here is that the great majority of supportive studies are flawed in some way. In several cases, the method used to assess beliefs may have indirectly influenced the responses given (e.g. Thouless,1956; Lefford,1946). In other cases, beliefs were not ascertained, they were simply assumed (e.g. Nehrke, 1972). Such problems are not peculiar to positive reports. Nevertheless, they severely undermine the argument put forward. Throughout Experiments 1 to 9, beliefs were ascertained from, and based on, the responses of a matched peer group which overcomes this type
of problem. A similar method was used by Kaufman & Goldstein (1967). Their study provides convincing and supportive evidence, but does not report a main effect of belief bias. The highly significant main effect of belief bias found in Experiments 1 to 3, therefore, represents the only body of experimental evidence to demonstrate a consistent and substantial main effect of belief bias on the evaluation task which could not have been induced or influenced by believability rating procedures.

Although it is important to emphasize the significance of these findings, we should be wary of assuming that these results reflect any true level or nature of the belief bias effect. In the syllogistic evaluation task the subject is presented with a single decision of acceptance or rejection. The 2x2 design, matching and cross matching logic and belief yields a strong effect of both factors. However, we must consider whether such effects were allowed, or, in fact, promoted by task demands. The dichotomous nature of response categories may make the decision appear more definitive than it actually is. Additionally, the explicit presentation of a single conclusion which is clearly believable or unbelievable may highlight this as a response criterion which might not otherwise spring to mind.

Using identical problem content, Experiments 4 and 5 showed no evidence of belief bias on the syllogistic construction task. There was no tendency to produce more believable than unbelievable conclusions. In fact, any tendency was in the opposite direction overall. There was, however, a larger number of believable than unbelievable conclusions on indeterminate problems, but this was equalled by the number of
nonpropositional conclusions produced. Of the three usual forms of the syllogistic task, the construction task places the least constraint on the subject, in terms of how a conclusion should be formulated. Unlike the evaluation task, no specific proposition is singled out for consideration. Unlike evaluation and multiple choice tasks, no specific order of terms is demanded. As a consequence of this unrestricted format, Experiments 4 and 5 discovered the presence of a hitherto undetected directionality preference in conclusion production. The tendency to place subset before superset is indicative of some form of semantically based problem analysis. The tendency demonstrates a sensitivity to set-size which, in this case, can only be divined by reference to real life knowledge of the terms used. Clearly, people "go beyond the information given" with the syllogistic construction task, but do not show belief bias when dealing with this type of problem content.

The only other study to examine belief bias on the syllogistic construction task was conducted by Oakhill & Johnson-Laird (1985a). Their findings suggest that there are, in fact, effects of belief bias when subjects produce conclusions of their own. The belief bias effect obtained was, however, peculiar to problems whose logical or dominant erroneous conclusion was definitionally as opposed to empirically false. Hence, the belief bias effect was selectively influenced by the nature of the belief in question. Since the beliefs addressed in Experiments 4 and 5 were largely empirical, we can resolve the apparent conflict between the two studies. The differential results may be attributable to differences in problem content. The implications
of this are discussed in the next section.

If we examine Oakhill & Johnson-Laird's results in more
detail we find that the effect of belief was more marked for
indeterminate than determinate premise pairs. In their second
experiment, which contained premise pairs which were structurally
similar to those used in Experiments 4 and 5, 89% of acceptable
erroneous conclusions were believable and the percentage of
unbelievable errors was only slightly higher when the correct
conclusion was unbelievable (6.8%), than when it was believable
(4.5%). Interestingly, when the logically correct conclusion was
unbelievable, subjects reacted to this conflict by selecting the
incorrect nonpropositional conclusion. This type of response
pattern was also noted by Revlin, Leirer, Yopp & Yopp (1980) in
their multiple choice task study. Common to both studies was the
finding that belief bias interacted with the nature of the belief
in question. Definitionally false conclusions were less likely to
be drawn or selected than empirically false conclusions.

Experiments 1 to 5 have demonstrated a clear-cut and
consistent belief bias effect, which appears to operate
selectively on different types of syllogistic task. On the
evaluation task, the effect is strong and replicable. On the
construction task the effect is absent. The implications of these
findings are clear. Since the empirical belief bias effect
appears to operate selectively on the evaluation, but not the
construction task it seems that belief bias may be induced or
inhibited by the manner in which a reasoning task is presented.
In the light of these results and those reported by Oakhill &
Johnson-Laird, two important factors appear to be a) the nature
of the belief and b) the nature of the reasoning task. The further implications of such findings are many and various. At this point we will focus on their usefulness in resolving the apparent conflict between the evaluation and multiple choice task findings outlined in Chapter 4.

Studies of the multiple choice task provide the largest body of negative and contradictory findings (e.g. Henle & Michael, 1956; Gordon, 1954; Madrzycki, 1978; Revlin, Leirer, Yopp & Yopp, 1980). To recap, briefly, Henle & Michael found no direct relationship between the conclusions selected and belief ratings. Similarly, Gordon found that only 11 out of 28 subjects selected conclusions which were consistent with their beliefs. Madrzycki's subjects tended to avoid the logically correct, believable, conclusion. In contrast to these findings, Revlin et al. found that errors increased when logic and belief conflicted. However, less than half the errors made were consistent with beliefs (47.8% on determinate and 4.3% on indeterminate syllogisms). In the light of the foregoing discussion we may argue that there are certain features inherent in the multiple choice task format which explain some of the variation in the degree and characterisation of belief bias in these studies.

8.2 What is the nature of belief bias?

Several studies of belief bias have reported a logic x belief interaction, hence it may be inappropriate to consider belief bias as an isolated main effect. The theoretical explanations so far advanced appear to overlook this point, possibly because they have neglected to clarify the exact nature
of belief bias. Belief bias can be examined in a number of ways, but a complete characterisation of the effect can only be obtained by employing properly rated, neutral comparison problems. No previous studies have used this method in a carefully controlled fashion. Furthermore, the majority of studies in this area of research have focused exclusively upon one response measure; selection, evaluation or production rates for various types of conclusion. One or two studies have collected protocol data in addition to this. On the basis of this type of qualitative analysis, Henle (1962) specifically has claimed that belief bias does not necessarily reflect an inability to perform logical inference. Instead, she has argued that the apparently non-logical tendency is based on problem misconception or misinterpretation.

The following sections attempt to clarify the nature of belief bias. They consider a) differences between responses to neutral and belief based problems, b) the interpretation and usefulness of protocol data and c) the implications of the logic x belief interaction.

8.3 Comparisons between neutral problems and those with obvious truth status.

Experiment 6 attempted to establish the nature of belief bias. A syllogistic evaluation task was employed. The mood and structural composition of the problems used were identical to those of Experiment 2. However, in this experiment, it was important to ensure that both believable and unbelievable conclusions were rated at extreme poles of the believability scale. In Experiments 1 to 3, unbelievable conclusions had not
received the extreme ratings required. Hence, new contents were introduced to achieve this end. In order to distinguish between a bilateral and unilateral belief bias it was necessary to include neutral problems for purposes of comparison. An initial attempt to obtain neutral ratings for the conclusions to these problems was thwarted by subjects judging them to be 'slightly believable', on average. Eventually, more permissible statements were constructed; although some were still rated as marginally believable. This should be borne in mind in interpreting the results. Sign tests on valids and invalids, comparing responses to neutral problems with those to believable and unbelievable showed a significant difference between performance on neutral and unbelievable problems, but only on invalids. In interpreting these results, the reader may find it useful to refer back to Table 7.1 and examine the top half, which relates to Standard instructions.

There is an interesting trend here. On invalid problems, neither believable nor unbelievable problems seem to be treated in the same way as neutrals. In this respect, there may be a bilateral influence of belief bias; although the distinction between believables and neutrals does not reach significance. This may be due to the marginally positive belief associated with neutral conclusions.

In terms of the overall percentage correct, performance on unbelievables is better than on neutrals which, in turn, is better than on believables. On this basis, it is unwarranted to assume that problem analysis is aided by disbelief; since the overall superiority is attributable to performance on the
Invalid-Unbelievable problem. In this case, belief dictates the logically correct response. However, if we disregard neutrals for the moment and examine conflict problems (where logic and belief disagree) we find that errors of acceptance are twice as frequent as errors of rejection. This pattern of errors is similar to that observed by Kaufman & Goldstein (1967). Because of the complex interaction effects found by Kaufman & Goldstein it is not possible to draw general conclusions about errors. However, there are similarities between their data and those of Experiment 6. We can sum up this similarity between studies by saying that subjects more readily accept than reject a conclusion in accordance with belief. Hence, although subjects may approach both believable and unbelievable problems in a different way to neutrals, the response patterns may not necessarily reflect equal and opposite tendencies.

8.4 The interpretation and implications of protocol data.

With the addition of protocol data, collected in Experiments 1 to 3, we are fortunate in being able to examine performance from another perspective. The analysis of response frequencies found in Experiments 1 to 3 does not provide us with a complete picture of underlying processes. The protocol data collected in Experiment 3 provided useful, if not conclusive evidence of different modes of reasoning. Each appeared to be associated with belief bias to differing degrees. From the analysis of concurrent verbalisations, protocols were found to fall into three distinct categories. About 40% of protocols appeared to focus on the conclusion alone, mentioning neither of the logical premises.
These protocols were associated with the highest level of belief-based conclusions. A second type of protocol mentioned the conclusion first and then went on to consider one or both premises. About one quarter of the protocols collected fell into this category. This type of protocol showed some sensitivity to logic. In this case, beliefs were more likely to influence the decision if the conclusion was invalid. A similar frequency of protocols fell into the third category which mentioned one or both premises first and then the conclusion. This type of protocol was associated with the greatest number of logically correct responses and showed least belief bias of all. This pattern of protocols, showing three distinct classes, suggests that belief bias may be exerted at different levels. The first, 'conclusion only', group of protocols may be associated with what Henle (1962) has termed 'rejection of the logical task'. They appear to reflect a direct, belief based evaluation of the conclusion. However, this only accounts for a relatively small proportion of responses. At the other extreme, very few subjects consistently produced logically correct responses. The majority of responses reflected a sensitivity to both logic and belief.

Protocol analyses used by Henle (1962) led her to postulate various types of error which influenced problem representation. On this basis, she argued that logical errors, reflecting beliefs, may be attributable to encoding failures, rather than any non-logical bias. Specifically, she postulated three types of error which occurred before or during the encoding stage: (i) Restatement of premises so that the intended meaning is changed, (ii) Omission of a premise, and (iii) Slipping in of additional
premises. In Chapter 6 example protocols, taken from Experiments 2 and 3, showed evidence of such features. Whilst it would be untrue to deny their occurrence, we should note that these protocols were frequently associated with correct as well as with erroneous responses. Hence, it is inconsistent to argue, as Henle does, that such protocols are indicative of erroneous encoding, yet at the same time contend that any correct response was derived logically. Furthermore, we must bear in mind that Henle collected protocols retrospectively. It is, therefore, possible that the verbalisations collected were rationalisations.

Wason & Evans (1975) have demonstrated that people will tailor an account of their reasoning on a particular problem to suit the conclusion given. The concurrent verbalisations collected in Experiments 2 and 3 are arguably less likely to represent rationalisations, but even in this case, they can only suggest that different modes of reasoning are associated with different types of response. We cannot infer any causal relationship between mode of reasoning implied and belief bias. Certainly, it seems that belief bias is more likely to occur when 'backward' (conclusion to premise) reasoning is suggested. The failure to establish any belief bias effect on the syllogistic construction tasks, used in Experiments 4 and 5, tends to support the notion that belief bias may be triggered by the presence of a conclusion having obvious truth status. When the premises alone are presented, subjects may be more inclined to process the information in a forward, premise to conclusion, fashion. According to protocols, this mode of reasoning is associated with least belief bias of all. This does not imply that belief bias
could not occur on the syllogistic construction task. Indeed, Oakhill & Johnson-Laird (1985a) have demonstrated this, but principally with definitively false conclusions. Perhaps the conclusion is highlighted in such a case. Although belief bias appears to be associated with conclusion-based, backward reasoning, this does not, in itself, explain why or how belief exerts its influence. Having said this, any explanation of belief bias which can also account for this tendency would be all the more convincing.

8.5 The logic x belief interaction.

We can see from figure 8.5.1 that the logic x belief interaction can be characterised in two ways. Although it is significant on both valid and invalid problems here, the effect of belief is clearly stronger on invalids. Alternatively, we find that logic has a more marked effect on unbelievable problems. The conflict problems indicate that errors of acceptance, in accordance with belief, are more frequent than errors of rejection. These are 70% and 44% respectively. The nature of the conflict is different in each case, of course. With the Valid-Unbelievable problem, a definite and necessary conclusion has to be rejected if the response is to accord with belief. With the Invalid-Believable problem, the conclusion offered is not definitely logically incorrect; it is possible, but not necessary. The weighting given to logic and belief, in each case, may vary with the logical status of the conclusion. Similar differential weightings of logic and belief have been shown with other paradigms. Revlin, Leirer, Yopp & Yopp (1980) and Oakhill &
Figure 3.5.1
Graphical illustration of the logic x belief interaction for data summed over experiments 1 to 3. a and b show the interaction from two different viewpoints.

Key:
B+ believable
B- unbelievable
V+ valid
V- invalid
Johnson-Laird (1985a) have reported this characteristic logic x belief interaction; first documented by Kaufman & Goldstein (1967). The two latter studies show interactions which closely resemble that shown in figure 8.5.1 which represents the logic x belief interaction for data summed over Experiments 1 to 3.

The logic x belief interaction was consistent across Experiments 1 to 3 (although it failed to reach significance in one instance). This, together with its generalisability across paradigms, suggests that this is a robust effect which is not due to the idiosyncracies of a particular task format. Although, in the first three experiments, belief bias is significant on both valid and invalid problems, this is not the case in Experiments 6 to 9. For these experiments there is no significant belief bias effect on valid problems. If we consider a) the differing nature of the subject population and b) established patterns of performance associated with subjects of differing levels of education and critical ability, there appears to be a reasonable explanation for the differences across blocks of experiments.

There is a two year gap between the first and last block of experiments in this series of nine. The examination entry requirements for these students rose considerably over this period. The educational level of the two populations is, therefore, different. In the belief bias literature, there are frequent references to factors such as educational level and critical ability. A consistent finding is that highly educated subjects and those with good critical ability show the least belief bias of all (Thouless, 1959; Feather, 1964,1967; Nehrke; 1972). If we extrapolate from these findings to those of
Experiments 6 to 9, it is reasonable to argue that those subjects with higher A'Level grades may give less weight to belief generally; resulting in the loss of belief bias on valid, but not invalid syllogisms. Since we cannot view belief bias and logic as discrete effects, we should reject the conventional idea that belief bias is simply a tendency to accept or reject according to belief. Although protocols suggest that some decisions are based solely on the believability of the conclusion, a large proportion show a sensitivity to logic also. However, depending on the nature of this conflict, each appears to be given a different weighting.

On the basis of experiments 1 to 3, three explanations of the logic x belief interaction were put forward in Chapter 7. They were tested in Experiments 6 to 9. The first explanation to be tested concerned the asymmetrical believability ratings which were given for believable and unbelievable statements. Since, unbelievable statements had not received an extreme rating on the believability scale, it was hypothesized that the conflict arising from a Valid-Unbelievable problem was less strong, in terms of belief than that arising from an Invalid-Believable problem. By replacing these unbelievable statements with those that had been rated at the extreme end of the believability scale, the balance could be restored for each conflict problem. The strengths of the conflicting beliefs should be equivalent. On the basis of this hypothesis, no logic x belief interaction was predicted when believable and unbelievable ratings were of equivalent strengths. Contrary to this prediction, the logic x belief interaction did not disappear with the new problem.
content. It was highly significant. In terms of theoretical generalisability, this explanation was the least appealing, since it seemed unlikely that other studies had employed believable and unbelievable statements with similarly asymmetrical ratings. Two further explanations were put forward: the 'misinterpreted necessity model' and the 'selective scrutiny model'. Both had theoretical generalisability and were based on empirical evidence drawn from the decision making literature. We will consider each model separately.

**The Misinterpreted Necessity Model.**

Dickstein (1980,1981) has put forward evidence which suggests that some subjects do not interpret the concept of necessity in a strictly logical sense. He claims that people feel warranted to accept a conclusion if it is plausible and does not contravene the premises. In other words, people do not appreciate that a conclusion can only be accepted as valid if it necessarily follows from the premises. According to this view, errors arise because of a misinterpretation rather than a non-logical bias. To recap, the reasoning procedure postulated by the misinterpreted necessity model is shown in figure 8.5.2. This type of explanation has certain features in common with the rationalist approach. Just as Revlin has claimed that the illogical component of his conversion model can be removed by conversion blocking, Dickstein has claimed that logical errors will be considerably reduced by clarifying the logical meaning of necessity. Hence, the mechanism of inference need not be illogical or non-logical in itself. Belief bias is simply due to a misinterpretation at
Figure 8.5.2

*Schematic diagram of the misinterpreted necessity model*

Does conclusion contravene premises?

- No
- Yes

Is it valid?

- Yes
- No

Is it believable?

- Yes
- No

Accept conclusion

Reject conclusion
the final stage of decision making.

Table 8.5.1

Percentage frequency of conclusions accepted in Experiments 6 to 8, divided according to type of instructions.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Valid</th>
<th>Invalid</th>
<th>Standard</th>
<th>Augmented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B+</td>
<td>B-</td>
<td>B+</td>
<td>B-</td>
</tr>
<tr>
<td>Experiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>75</td>
<td>75</td>
<td>71</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>84</td>
<td>80</td>
<td>81</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>16</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>--</td>
<td>--</td>
<td>72</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>35</td>
<td>8</td>
</tr>
</tbody>
</table>

Experiments 6 to 8 attempted to test the misinterpreted necessity model by manipulating instructions. Each subject received either Standard or Augmented instructions. Standard instructions were based on the instructions given in Experiment 3. Augmented instructions differed from these in one respect only; they included an explanation of logical necessity, based on that given by Dickstein (1981) in his instructions. A summary of
response profiles is shown in table 8.5.1.

In Experiment 6, as predicted, the effects of logic and belief were significant for the Standard instruction group. There was also a highly significant logic x belief interaction. There was no significant effect of belief on valid problems this time. For the Augmented instruction group, both the belief bias effect and the logic x belief interaction disappeared. However, comparisons between groups revealed significant differences for belief indices only. The failure to establish a significant difference between interaction scores was probably due to a slight tendency towards interaction under Augmented instructions. Interestingly, the nature of this interaction was uncharacteristic of that detected in earlier experiments. As usual, there was a slightly larger number of believable conclusions accepted on invalid problems, but on valid problems, if anything, this tendency was reversed. Two reasons for the unpredicted behaviour of belief bias on valids were examined and ruled out in Experiment 7. Neither the new problem content nor the within subject inclusion of neutral problems appeared to be responsible for this pattern of results.

Let us first consider the response pattern on invalid problems since the misinterpreted necessity model is chiefly concerned with errors on such problems. Specifically, it predicts that belief should not influence responses to such problems under Augmented instructions. It is true that, in Experiment 6, the acceptance rate for Invalid-Believable problems was substantially reduced under Augmented instructions. However, a reduction of this magnitude was not evident in Experiments 7 and 8, which had
attempted to replicate this finding. In both experiments, there was a significant belief bias effect on invalid problems under Augmented instructions. There were no differences between interaction indices for the Standard and Augmented instruction groups. These results present a serious problem for the misinterpreted necessity model. The predicted, dramatic, fall in Invalid-Believable acceptance rates was observed in only one out of three experiments.

If we now examine response profiles on valid problems the misinterpreted necessity model encounters further difficulties. According to this model, Augmented instructions should have no effect on the responses to valid problems. If we compare the pattern of responses across instruction groups, along with the fall in acceptance rates for Invalid-Believable problems, we find that the level of acceptance for Valid-Believable also falls although not so markedly. Clearly, Augmented instructions produced a more diverse effect than that predicted by the misinterpreted necessity model, yet the overall influence was not unsystematic. Subjects now appeared to be questioning believable conclusions, resulting in a noticeable fall in acceptance rates for Invalid-Believable problems and a slight fall for Valid-Believable problems. (We would not ordinarily expect the acceptance rate for Valid-Believable problems to fall if they were being subjected to analysis. However, this is in keeping with the idea that logical performance was formerly artificially augmented by belief).

If the logic x belief interaction were created by a simple misunderstanding of logical necessity, the clarification of this
concept should be sufficient to remove the effect. Experiments 6 to 8 failed to obliterate the logic x belief interaction with Augmented instructions. Furthermore, the effect of Augmented instructions did not operate selectively on Invalid-Believable problems. In the light of this evidence, the misinterpreted necessity model was rejected. However, the problem of explaining why Augmented instructions had such effects remained. The problem appears to be resolved if an additional interpretation of Augmented instructions is considered.

In the introduction to Experiment 9, it was argued that some subjects could have interpreted the Augmented instructions as a subtle hint not to accept conclusions just because they were believable. Greater scrutiny of believable problems would lead to an overall increase in the number of correct responses on such problems. In theoretical terms, the results obtained are compatible with the predictions made by the selective scrutiny model. This model claims that subjects do not analyse believable problems. On unbelievable problems, the search for a falsifying alternative is initiated by pre-existing knowledge of an alternative, believable, conclusion. This belief prompts and guides further analysis. This does not occur on believable problems. Hence, the high level of correct responses on Valid-Believable problems would not necessarily reflect any logical analysis. This level is artificially high due to the coincidence of logic and belief. According to the selective scrutiny model, if subjects begin to analyse the validity of believable conclusions, as a result of instructions perhaps, we might expect a more marked improvement on Invalid-Believable than
on Valid-Believable problems. This would be because the logically correct response rate was previously augmented by belief on the latter problem type. However, we must acknowledge that this is a post hoc application of this model. Nevertheless, such an explanation becomes increasingly convincing in the light of Experiment 9.

In this experiment subjects were either given 'Complex' or 'Simple' instructions. Complex instructions were based on those given to accompany the syllogistic construction task of Experiments 4 and 5, but this time they were given with the syllogistic evaluation task and amended accordingly. Such instructions could be viewed as placing additional emphasis on logical structure by giving an abstract example and explaining that conclusions should be based on the structural relationship stated in the premises. The results of Experiment 9 provided convincing support for this argument. Simple instructions closely resembled Standard instructions. They were based on Complex instructions with the example syllogism and its explanation removed. The repetitive emphasis placed upon the logical task was also minimised. Under Simple instructions, there were highly significant effects of logic and belief and an interaction between the two, as predicted. However, under Complex instructions, the effect of belief fell just short of significance (P = 0.055, one-tailed test), and there was no logic x belief interaction. The logic x belief interaction disappeared for two reasons. Performance on Invalid-Believable problems improved, but performance on Invalid-Unbelievables deteriorated. This is illustrated in table 8.5.2, which views performance on
each of the four problem types in terms of percentage of conclusions accepted.

Table 8.5.2

Percentage frequency of conclusions accepted for each of the four problem types. (Experiment 9).

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>B+</td>
<td>B-</td>
<td>B+</td>
</tr>
<tr>
<td>Valid</td>
<td>81</td>
<td>69</td>
</tr>
<tr>
<td>Invalid</td>
<td>81</td>
<td>19</td>
</tr>
</tbody>
</table>

From table 8.5.2 it appears that emphasis on logical structure does not increase logically correct performance on all problem types. The effect appears to operate selectively on the Invalid-Believable and the Invalid-Unbelievable problem. In accordance with the selective scrutiny model, this differential effect suggests that there may be a dual effect of belief under Simple instructions. Believable problems appear to receive relatively little logical scrutiny, whereas unbelievable problems receive either greater than usual or belief directed scrutinization. The latter alternative appears most likely. If the presented conclusion is unbelievable, everyday experience could prompt the consideration of a falsifying alternative; which
is already available as a pre-existing belief. When the given conclusion is believable, such an alternative is unlikely to be so salient.

Table 8.5.3

Comparison of performance under Complex instructions (Experiment 9) with that on neutral problems (Experiment 6, Standard instructions). Figures shown represent percentage of conclusions accepted for each problem type.

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Neutral</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>B+</td>
<td>83</td>
<td>81</td>
</tr>
<tr>
<td>B-</td>
<td>63</td>
<td>31</td>
</tr>
</tbody>
</table>

This explanation of results has far reaching implications. Firstly, it alters the conventional view of belief bias. It implies that although positive belief may be detrimental to achieving a logically correct conclusion, a negative belief may improve the chances of a logically correct response. If we encourage subjects to disregard totally both positive and negative beliefs (as with Complex instructions), we find a fall in logically correct responses to unbelievable problems. This is
because we have taken away an aid to analysis. It is interesting to note that performance under Complex instructions then resembles that observed with neutral problem content in Experiments 6 and 8. This is possibly because there is no guidance provided by belief in either case (see Table 3.5.3).

Of the three models of the logic x belief interaction, the selective scrutiny model appears to be the most appropriate, by virtue of the fact that it is the only account to remain intact. In theoretical terms also, the selective scrutiny model is the most appealing. It is consistent with the qualitative as well as quantitative data collected in Experiments 1 to 3. It is generalisable across different inference tasks. Moreover, it is compatible with current approaches to human inference. Notably, the selective scrutiny model is related to models of belief bias effects put forward by Oakhill & Johnson-Laird (1985a).

Chapter 9 attempts to account for the results of Experiments 6 to 9 in terms of the selective scrutiny model. It goes on to compare this model with other theories of belief bias and places it within the context of more general theories of human inference.
Chapter 9

The selective scrutiny model of belief bias. An integration of empirical evidence and a consideration of wider issues.

9.1 An attempt to resolve conflict ........................................... 260
9.2 Unbelievable statements - the wider domain ....................... 265
9.3 Selective scrutiny and theories of belief bias. ...................... 267
9.4 Selective scrutiny and theories of syllogistic reasoning. ........ 272
9.5 General theoretical issues .............................................. 276
9.6 Selective attention and everyday decision making. ............... 278
9.7 Belief bias and future research ..................................... 280
9.1 An attempt to resolve conflict.

Studies of belief bias vary a great deal. Several yield conflicting results, but experimental design varies to such an extent that we cannot pin down these conflicting results to any one specific factor. Idiosyncratic characteristics of specific studies have already been criticised. The aim of this section is to bring together conflicting findings, viewing them from a single theoretical standpoint. Figure 9.1.1 presents a schematic diagram of the selective scrutiny model.

Firstly, let us examine the findings relating to Experiments 1 to 9. From the nine experiments carried out here, we can highlight four potentially influential factors, which may help to reconcile conflicting elements in this area of research:

(i) Belief bias is stronger on invalid syllogisms. Although, in the first three experiments, belief bias is significant on both valid and invalid problems, this is not the case in Experiments 6 to 9. For these experiments there is no significant belief bias effect on valid problems.

(ii) Belief bias appears to be sensitive to paradigm. There was a significant effect of empirical beliefs on the evaluation, but not the construction task. We know that belief bias can occur on the construction task, but this appears to be largely restricted to definitional beliefs (Oakhill & Johnson-Laird, 1985a).

It follows from (ii) that:

(iii) Belief bias is sensitive to the type of belief addressed. Indeed, there may well be an interaction between (ii) and (iii).

Finally, we know that (iv) the logic x belief interaction is sensitive to instructional variation. Instructions which
Figure 9.1.1

Schematic diagram of the selective scrutiny model.

Is conclusion believable?

Yes

Is there a falsifying alternative which is consistent with the premises?

No

Accept conclusion

No

Reject conclusion

Yes
emphasise logical structure are associated with a levelling out of the differential effects belief on invalids.

Experiments 1 to 9 and their theoretical interpretation may help us to unravel problems of conflict between existing studies. Bearing in mind points (i) to (iv) above we will now go on to consider this possibility.

There appear to be five studies in particular which report an unconventional belief bias effect, if any. The majority employed the multiple choice task (Henle & Michael, 1956; Madrzycki, 1978; Gordon, 1953; Revlin, Leirer, Yopp & Yopp, 1980). This common element may be significant. Theoretically, we might expect less belief bias on this type of task, because attention would be more evenly distributed over conclusions due to the presentation format used. Frase's (1966a,b; 1968) studies are the only evaluation task experiments which failed to show any conventional belief bias effect. In these experiments the conclusions contained pairs of terms with differing levels of incompatibility; which appears to be correlated with levels of believability. Examination of the conclusions presented suggests that many were rather novel or obscure and ponderable. As such they are uncharacteristic of items used in this area of research. This leads us to question the psychological equivalence of Frase's tasks and those usually presented in studies of belief bias. We, therefore, concentrate on the four multiple choice task studies.

Notably, two of these studies employed a preponderance of valid syllogisms. We know empirically and predict theoretically that belief bias will be less likely to appear on this type of
problem. None of the four studies gives details of the instructions used. Henle & Michael placed much emphasis on the importance of instructions. Unfortunately, they have since thrown theirs away (M. Henle, personal communication). Therefore, we cannot tell if the instructions used acted to emphasise logical structure in any way. There are, of course, other means by which this can be achieved. Madryzcki tested only those subjects who had completed a course in logic beforehand. These two experiments have two interesting features. Henle & Michael's subjects showed no belief bias, but performance was unimpressive until they were instructed in the use of diagrams to aid them in their task. This may have acted to change the nature of the reasoning process. Memory load would be reduced and thus enable a systematic and possibly exhaustive search for alternative conclusions. Hence, their findings are consistent with the selective scrutiny model. Interestingly, whilst Madryzcki's subjects showed no preference for belief-consistent conclusions, they were more likely to select the logically correct conclusion if the dominant (normally favoured) conclusion was unbelievable. This is entirely consistent with the selective scrutiny model's predictions. Hence, even if subjects are conversent with logic, negative belief may still aid them in their analysis; by directing their attention immediately to a potentially falsifying instance.

The aid to 'logical analysis' provided by a negative belief is evident across many studies. Table 9.1.1 illustrates this pattern.
Table 9.1.1

Studies of belief bias classified according to logical performance on unbelievable problems.

<table>
<thead>
<tr>
<th>SUPERIOR</th>
<th>INFERIOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janis &amp; Frick</td>
<td>Frase</td>
</tr>
<tr>
<td>Kaufman &amp; Goldstein</td>
<td></td>
</tr>
<tr>
<td>Henle &amp; Michael</td>
<td></td>
</tr>
<tr>
<td>Madryzcki</td>
<td></td>
</tr>
<tr>
<td>Revlin et al.</td>
<td></td>
</tr>
<tr>
<td>Oakhill &amp; Johnson-Laird</td>
<td></td>
</tr>
</tbody>
</table>

Studies which are not listed in the above table provide insufficient evidence for us to ascertain performance on believable and unbelievable problems separately. Studies listed in the first column suggest an interesting response pattern. Although, in some cases a conventional belief bias effect was not observed, in other aspects there are striking similarities across studies. In almost every case there was an overall tendency for subjects to make more errors of acceptance than errors of rejection. Where such differences do not exist, we find that the nature of errors on believable and unbelievable problems (i.e. those which have a dominant response in abstract form which would favour one or the other, or those whose presented conclusion constitutes one or the other) are noticeably different. Clearly,
this evidence suggests that belief bias is not a symmetrical effect. Just as the selective scrutiny model predicts, there are fewer errors on unbelievable problems. Where errors do exist, they are not generally indicative of blind rejection.

9.2 Unbelievable statements - the wider domain.

The pattern of results reported above is not peculiar to syllogistic reasoning. Similar tendencies are evident in conditional reasoning tasks. Using Wason's selection task, (where the logically appropriate aim is to falsify the rule presented), Van Duyne (1976) claimed that subjects performed significantly better with rules which they believed to be contingently rather than necessarily true. In a later experiment, Pollard & Evans (1981) criticise Van Duyne's analysis. However, in a follow up study, based on this, their results also suggested that performance was better when the rule presented was empirically false as opposed to empirically true. In a later study, however, Pollard & Evans (1983) report somewhat different effects. Logical performance appeared to be better on believable rules. There was evidence to suggest that subjects subjected unbelievable rules to greater scrutiny, but this did not aid them in their final decision. Hence, it seems that unbelievable statements were scrutinised, but that the tests employed were not guided by belief. The rules used referred to general tendencies rather than facts. Hence, belief may have initiated scrutiny of the evidence, but the belief or everyday knowledge in question may not have been so strongly established as to cue selective consideration of the counterexample. Hence, the selective scrutiny model can
We do not need to postulate the uncovering of any logical competence to explain superior performance on unbelievable problems. Clearly, in terms of the selective scrutiny model an explanation is obvious. In cases where the rule is empirically false, a more salient alternative corresponds to the potentially falsifying case - selection of the $\bar{Q}$ card. Past experience cues the logically correct selection. However, from Chapter 2, Section 3 we can see that a simple memory cueing explanation is insufficient in explaining related thematic effects. The facilitatory effect of thematic materials was apparently bound up with the so called 'detective set'; which is characterisable as a tendency towards suspicion or scrutiny. Parallels between this explanation and the characterisation of belief bias presented here are clear.

Suspicion or scrutiny involves viewing information from a biased standpoint; one is looking for flaws because they are, in a sense, expected. Such a tendency has been demonstrated in a study by Lord, Ross & Lepper (1979). In this study, subjects attacked and criticised studies which opposed their point of view on the question of hanging. Studies which supported their stance were handled with "kid gloves". Other effects may also be explained by appealing to this tendency. Many linguistic effects are apparently attributable to presuppositions which may act to channel information processing in a specific direction. Presentation effects, such as those discussed in Chapter 2 section 3, may create expectancies in a similar way. The manner in which a question is phrased may alter the way we treat the
information presented. Examples are readily evident in the area of propositional reasoning. In the Wason selection task, for example, subjects can be asked to test the given rule in various ways. The traditional question is usually a request to 'test the truth or falsity of the rule'. Recently, subjects have been asked to decide if someone is 'correct' or 'telling the truth', (Reich & Ruth, 1982). In this case, the majority of responses appeared to reflect either verification or falsification strategies instead of the usual matching bias. Clearly, the way in which the question was framed is likely to have influenced the selection of cards here.

In the light of such similarities it is unparsimonious to view such effects as the reflection of disparate processes. Most evident is that belief bias can be viewed as another instance of the 'detective set'. However, certain presentation and linguistic effects may also be related. In each case we can see a channelling of thought; created on the basis of presuppositions. There are undoubtedly other ways of directing thought by increasing the likelihood of specific alternative possibilities.

9.3 Selective scrutiny and theories of belief bias.

As we have already noted, theories of belief bias are not prolific. Explanations of belief bias appear to fall into three distinct camps. Firstly, we have Henle's (1962) rationalist approach. According to this view, rationality is not challenged by the existence of systematic, belief-based errors. Belief bias is seen to be the result of misunderstanding or task rejection.
As such, it is not characterised as an instance of non-logical reasoning. Undoubtedly, there are some cases in which subjects reject the logical task. Protocols suggest that this is so. However, a large proportion of responses appear to result from a genuine attempt to solve a logical problem. The only specific theory of misinterpretation is Revlin's adapted Conversion theory. Since conversion of premises was controlled in all but two of the experiments reported here we can rule out this explanation. Another possible source of misinterpretation was also tested in Chapter 7. However, the misinterpretation of necessity was found to be an inadequate explanation of belief bias on this task (see Chapter 8).

We are, therefore, left with two possibilities; either belief influences reasoning, or it replaces reasoning. The idea that belief, rather than logic, forms the basis of responses was put forward by Pollard & Evans (1981). This explanation was originally applied to the Wason selection task, but it is easily and directly applied to syllogistic reasoning. Specifically, Pollard & Evans claimed that responses to problems involving a priori beliefs were mediated by pre-existing associations between the terms mentioned. With believable statements, the constituent terms would be strongly associated. Conversely, with unbelievable statements, there would only be a weak association. Hence, belief bias could be the result of retrieval processes rather than any affective influence on reasoning. The distinction between the consequences of a cognitive mechanism and an affective or motivational disruption is an important one. As Pollard & Evans point out, beliefs evolve from past experience. However, past
experience also leads to other forms of information categorisation which do not directly involve any affective component. The linguistic presuppositions noted by Geis & Zwicky (1971) and Evans & Newstead (1977) are examples of this. Hence, it may be more fitting to view belief bias as a consequence of prior associations than any desire to maintain one's belief system. The associational interpretation implies that belief bias is a passive response involving no active analysis of new information. If a statement contains terms which are strongly associated it will probably be accepted. This implies that belief bias will result in equal and opposite tendencies for believable and unbelievable problems. However, the results of Experiments 1 to 9 and the findings reported in other studies (e.g. Revlin, Leirer, Yopp & Yopp, 1980; Oakhill & Johnson-Laird, 1985a) suggest that this is not necessarily the case. Subjects do not reject one conclusion in order to accept a conclusion with stronger associations. This point is taken up again with reference to Pollard's Availability theory in section 9.5.

An alternative approach, much criticised by Pollard & Evans, was put forward by Van Duyne (1976). The theory of cognitive self reinforcement stresses the role played by motivational factors in reasoning. Specifically, Van Duyne claims that, if a rule is interpreted as necessarily true, subjects do not expect to find counterexamples. If a rule is interpreted as contingently true, the subject expects alternatives to exist. The subject then attempts to confirm this hypothesis or expectation. Hence, belief bias is interpreted as a tendency to confirm one's expectations. This tendency is designed to gain reward or cognitive self
reinforcement. The study reported by Van Duyne (1976) has certain flaws, which are pointed out by Pollard & Evans (1981) who argue that a motivational explanation is unparsimonious and explanatorily inadequate. In some respects, parts of the selective scrutiny model could be confused with elements of the cognitive self reinforcement theory. The two types of explanation should be clearly distinguished. Firstly, the selective scrutiny model does not attribute belief bias to any motivational influence on reasoning. Secondly, although this model claims that subjects seek counter examples if they believe that they are likely to exist, this does not imply that the subject is purely aiming to verify. He is aiming to falsify via verification of an alternative. Prior belief tells the subject that the existence of certain evidence can be used to falsify the conclusion or rule under consideration.

Hence, prior knowledge may help people to recognise the usefulness or relevance of a particular piece of information. This may be a crucial factor which underlies logical performance. Experiments by Oakhill & Johnson-Laird (1985(b)) support this idea. They suggest that appreciating what counts as a counterexample is a prerequisite which is often absent from performance in reasoning experiments. Oakhill & Johnson-Laird note that this recognition may develop with practice at a task or, alternatively, past experience may cue consideration of logically pertinent information. There are various explanations for this apparent failure to consider counterexamples. One explanation is that the subjects' working memory is not large enough to store and manipulate the required amount of
information. Alternatively, it could be that, subjects either do not a) appreciate the need or b) understand how to falsify in the unusual context of a formal reasoning task.

The idea that believability may influence the search process is directly related to one of two explanations put forward by Oakhill & Johnson-Laird (1985a). Studying syllogistic reasoning and applying a mental models approach (Johnson-Laird, e.g. 1983), they argue that belief bias may exert an influence in two possible ways. One explanation is that the believability of a putative conclusion influences the inferential process by terminating any search for counterexamples. If the subject arrives at a believable conclusion, there is no determined search for counterexamples. Alternatively, if the subject arrives at an unbelievable conclusion, it will be rejected in favour of a modified version, or the claim that no valid conclusion follows. Hence, the search for counterexamples is again suspended, but for a different reason. Although this explanation fits data from the multiple choice and construction tasks, it is difficult to see how it could account for behaviour on the evaluation task; where subjects do not totally reject unbelievable conclusions, on the whole. This difference in behaviour could be attributable to either paradigm or content. This observation is directly relevant to the second explanation put forward. In this case, it is suggested that subjects arrive at an unbelievable conclusion, but stating that conclusion in explicit verbal terms may engage beliefs which influence the likelihood that the conclusion will be expressed. For instance, subjects may feel wary of suggesting that a semantically absurd relationship holds between two terms.
The experiments reported by Oakhill & Johnson-Laird cannot distinguish between these two alternative explanations. The problem here is that neither explanation appears to account for performance on the syllogistic evaluation task. The first account, which attributes belief bias to a suspension of the search process, predicts rejection of unbelievable problems with equal strength on both valid and invalid problems. The second account makes an identical prediction.

How can we account for the observed differences across paradigm and content. One possibility is that a percentage of people prefer not to make semantically absurd statements, but feel less inhibited if a) this conclusion is the only one to be explicitly stated in the problem - and hence they are not the only ones to make such an assertion and b) the belief which is contradicted is not strongly held. If we combine this demand characteristic with the selective scrutiny model we could explain why logical performance is so impressive on unbelievable problems on the evaluation task and why the accuracy level on determinate, unbelievable problems is not so impressive on the multiple choice and construction tasks.

9.4 Selective scrutiny and theories of syllogistic reasoning.

Chapter 3 considered theories of syllogistic reasoning. These were subdivided according to the stage at which reasoning errors were claimed to occur:
Errors in premise interpretation

The conversion theories proposed by Chapman & Chapman (1959) and later by Revlis (1975b) and Revlin et al. (1980) cannot account for the data obtained, as we have already noted, there was no possibility of illegal premise conversion in the majority of the experiments carried out in the series reported here. The remaining theories; Caplan's (1981) Play it Safe model and Ekberg & Lopes' (1979) model of Natural Deduction, assume that errors reflect peculiarities in quantifier interpretation (within the premises). Clearly, this type of explanation does not coincide with the backward, conclusion to premise and hence conclusion based process detected here.

Errors in combining premises.

The set theoretic (e.g. Erickson, 1974) and Transitive Chain (e.g. Guyote & Sternberg, 1981) models of syllogistic reasoning both attribute a large proportion of error to a failure to consider all possible premise combinations. Such an explanation is consistent with the selective scrutiny approach. However, Erickson's model assumes that only one interpretation of each premise is randomly combined with the one interpretation of the other premise. Because the combination is random it cannot explain the systematic error pattern observed here. Guyote & Sternberg have tested the transitive chain model using thematic content. They claim that factual (believable) problems are associated with significantly more correct responses than counterfactual (unbelievable) and other types of content. This difference was attributed to a greater probability of combining
more than one premise representation in the case of factual problems. This claim is made despite the fact that there is no direct proof of differential frequency of premise combinations according to problem content. Clearly, this element of the theory is in direct conflict with the selective scrutiny model. There is, of course, a simple way to test this aspect of the transitive chain theory and that is to compare response times across problems with believable and unbelievable conclusions.

The two remaining theories in this category have several similar features. An important characteristic of both is that they can account for the well documented figural effect. Notably, the set theoretic and the transitive chain models fall down in this respect. It seems that the mental models explanation (Johnson-Laird, 1983) grew out of the earlier analogical theory (Johnson-Laird & Steedman, 1978). The mental models theory has been extensively developed and is not solely restricted to the area of syllogistic reasoning. A chief attraction of the mental models theory is its ability to account for logically correct responses as well as errors without resorting to abstract rules of inference. This theory claims that premises are represented as mental models which are integrated to form combined premise representations. On the basis of a syllogism by syllogism analysis, Johnson-Laird claims that no syllogism requires the construction of more than three integrated models. Accuracy is dependent upon (i) the number of models which need to be considered and (ii) the number of premise manipulations required to form easily integratable models. Erroneous response patterns appear to show that people rarely consider all possible
integrated models. Indeed, they tend to have preferences which determine which of a limited number of models is initially considered.

The assumptions made by the selective scrutiny model are consistent with a general mental models approach to reasoning. Errors are seen to reflect a limited capacity reasoning system. Specifically, they are due to the failure to consider alternatives. One difference between the mental models theory and the selective scrutiny model is the prediction of which integrated model will be considered first. This must be specified in order to predict particular error patterns. According to the selective scrutiny model, the construction of integrated models would be driven by the subjective nature of the claim made by the conclusion. This means that we are dealing with a 'backward' reasoning process.

Error at the conclusion translation stage.

Dickstein's directional processing theory is the only one to fit neatly into this category. Clearly, this theory rests crucially on the assumption that reasoning is a forward as opposed to a backward process. Errors are due to directional (A-C, C-A) inconsistencies between the entirely logical conclusion derived and the conclusion(s) presented. This view is clearly contrary to that expressed in the selective scrutiny model.

The remaining explanations, put forward by Newell (1981) and Fisher (1981), attribute errors to each stage of reasoning. Newell's application of the Problem Space explanation does not constitute a specific account of syllogistic reasoning performance. However, it sees syllogistic reasoning as goal
directed and influenced by set or past experience. Such an interpretation is consistent with the selective scrutiny model and constitutes a promising context in which to place more specific theories. It is considered as a general theory in the next section. In his investigation of syllogistic reasoning, Fisher has dissected the syllogistic task into discrete parts. As such, he seems to neglect the psychological difference between dealing with problems piece by piece, as opposed to a complete whole. In effect, this runs against the philosophy of the selective scrutiny model.

9.5 General theoretical issues

The existence of a belief bias effect has been viewed as a direct challenge by many supporters of the rationalist school. This, more than anything, seems to suggest that people are insensitive to the laws of logic. The proposed equivalence of human reasoning and formal logic has more recently been replaced by theories of natural logic or inference schemata. However, they still retain a competence / performance distinction. There are many problems with such views of reasoning, which were elaborated in detail in Chapter 1. Such problems have led to the criticism of logic based competence systems. However, some of the alternatives have been less specific. Cohen (1981) and Kyburg (1983) have presented particularly stimulating arguments in defending Man's rationality. Cohen claims that any response that has the backing of intuition and consensus is rational. Kyburg claims that people are quite right to evaluate conclusions according to their practical certainty - their likelihood of
being correct. He cites the confirmation bias, documented by Mynatt et al. (1977;1978) as a reasonable, rational approach. In terms of the selective scrutiny model, when reasoning, we act on cues or clues and confirmation bias could simply be another example of the channelling of thought.

Whether the selective scrutiny model presents Man as rational or not is a matter of debate. We must not be distracted by this issue and lose sight of a more tangible bone of contention. If reasoning is not based on logic, what is the mechanism of human inference? Those theories which reject the notion of a mental logic seem to have a unifying theme. Each views reasoning to be deep seated in experience. Frequently, analysis is short circuited, as a consequence, information is seldom examined completely. Theories which fall into this category vary in the extent to which analytical processes are involved. Perhaps the most extreme explanation has been put forward by Pollard (1982) who appears to claim that the response which first springs to mind will constitute the decision made. According to this view belief bias is due to believable statements being highly available and unbelievable statements being highly unavailable. Hence, we rarely observe the acceptance, selection or production of unbelievable conclusions. The availability hypothesis cannot explain the failure to obtain belief bias on the construction tasks of Experiments 4 and 5. Even if empirical beliefs are not as salient as some others (i.e. definitional) beliefs, they should certainly be more available than the unbelievable empirical alternative. The availability hypothesis also fails to explain the observed logic x belief
interaction, which suggests that believable conclusions may be analysed in a different way from their unbelievable counterparts. The availability hypothesis predicts a straight belief bias effect, since responses are produced passively subjects should not exhibit any sensitivity to logical structure.

The remaining explanations, put forward by Wason (1983), Evans (1984), Johnson-Laird (1983) and Newell (1981) each assume an element of analysis, but this analysis is largely determined by selective attention. The selective scrutiny model fits neatly into this type of theoretical approach.

9.6 **Selective attention and everyday decision making.**

Social psychologists have long been concerned with the manner in which we make social judgements. The expectancy-confirmation process is an important link in the chain leading from social perception to social action (Darley & Fazio, 1980; Snyder & Swann, 1978a). Research has identified two processes leading to confirmation of beliefs in social interaction. The first, labelled a "behavioural confirmation effect" is similar to Merton's (1948) self-fulfilling prophecy. The second process has been labelled a "cognitive confirmation effect". In this case people selectively interpret, attribute or recall information in ways which are consistent with their expectations (Duncan, 1976; Kelley, 1950; Langer & Abelson, 1974). In this way, people with different expectancies may witness an identical sequence of behaviour and still emerge with divergent expectancies confirmed. Recently, Darley & Gross (1983) have examined the process leading to the confirmation of
expectancies. With specific reference to stereotypes, which are arguably closely related to other types of attitudes and beliefs, they suggest that stereotypic information leads to the creation of hypotheses which are often tested in a biased fashion which leads to their false confirmation.

In a detailed review of the social judgement literature, Nisbett & Ross (1980) have documented a large number of cases in which subjects allow what they already know to influence the decisions they make. New and conflicting information, viewed in the light of prior beliefs, is often treated with scepticism and subjected to thorough and over-zealous scrutiny. Indeed, we have real life evidence of such behaviour in Garcia's (1981) troubling account of his repeatedly thwarted attempts to publish controversial findings in a psychological journal. Reviewers consistently rejected his evidence, apparently because it conflicted with their well established prior opinions on the matter. Evidence suggests that this bias may be overcome by the introduction of highlighting procedures. For instance, studies by Kaplan (e.g. Kaplan & Miller, 1978; Kaplan & Schersching, 1978) have reduced the effect of juror bias by enhancing the attention paid to relevant evidence.

Hypothesis confirming biases have also been reported in other areas of human decision making (e.g. Einhorn & Hogarth, 1978). The pervasive nature of this type of behaviour suggests that we should view it in terms of the information processing capabilities which an individual brings to bear on the many and various situations he encounters. Clearly, the distinction between 'Hot and Cold ' cognitions is unnecessary in this case,
since it would constitute an unparsimonious account of behaviour.

9.7 Belief bias and future research.

The foregoing discussion has aimed to clarify the issue of belief bias in syllogistic reasoning. Previous attempts to establish the nature and strength of belief bias have been unsatisfactory; mainly because of methodological flaws. Experiments 1 to 9 have suggested that belief bias is not a simple tendency to accept or reject conclusions according to their believability. The effect appears to be more subtle than this. As some previous studies have suggested, belief bias appears to interact with logic. Moreover, the belief bias effect is substantial when compared with that of logic, its strength is, indeed, sizeable enough to warrant concern.

The existence of belief bias presents a major threat to those who wish to protect the supposition that Man is inherently rational. Hence, rationalist theories have attributed this effect to task rejection or premise misunderstanding. In this way, the logical thought process goes unchallenged. The experiments reported here clearly demonstrate that, although some subjects may reject the logical task and some may also misunderstand the premises, such tendencies cannot account for the vast majority of responses observed.

Because of its title, belief bias has often been characterised as the distortion of reasoning by motivational forces. There is no direct evidence to suggest that belief bias has a motivational basis. Moreover, there are certain similarities between belief bias and error patterns on other
types of task. If response patterns on both abstract and thematic tasks can be similarly characterised, it is unnecessary to make any appeal to motivational substrates in explaining belief bias.

If belief bias does not arise from encoding failure, task rejection or the distortion of reasoning, we are left with an interesting alternative. Belief bias could be a direct consequence of the inferential processes which we bring to bear on new information. If this is the case, it is difficult to see how any theory which postulates formal inference rules could accommodate this effect. If we reject the notion of formal or natural inference rules, new avenues are opened. Current theoretical views claim that information analysis is seldom complete, and performance can be influenced by the channelling of attention. Indeed, many different types of error could result from this selectivity in human reasoning.

The selective scrutiny model is closely allied to this general approach. This model grew out of experiments on the syllogistic evaluation task, but it is equally applicable to other reasoning paradigms. Specifically, it claims that people selectively scrutinise conclusions which challenge prior beliefs. Believable claims are allowed to pass largely unchallenged. The exact mechanism underlying this procedure is, as yet, unspecified. It lies beyond the bounds of this particular research. The implications of this are clear. Whilst such a strategy is economical in terms of time and effort, it will lead to the maintenance of invalid beliefs unless they are challenged directly. Although belief bias does not seem to have unduly detrimental consequences for the individual, it is both
practically and theoretically important to discover how this limiting condition can be overcome. The selective scrutiny model predicts that this may be achieved by highlighting the variety of competing alternative possibilities to any one conclusion. This procedure has already received some attention in other areas of research. However, to the knowledge of this author, there is no existing research of a similar nature in the field of syllogistic reasoning.
REFERENCES

In R.S. Nickerson (ed) Attention and Performance VIII. 
Italian Journal of Psychology, 5, 245-249.
atmosphere and conversion interpretations of syllogistic 
reasoning errors. Journal of Experimental Psychology, 81, 
351-354.
Journal of Verbal Learning and Verbal Behavior, 21, 
595-620.
inferential problem as a function of stimulus materials. 
Quarterly Journal of Experimental Psychology, 26, 480-488.
Braine, M.D.S. (1978) On the relation between the natural logic 
of reasoning and standard logic. Psychological Review, 85, 
1-21.
propositions. Cognition, 6, 55-77.
Budner, S. (1962) Intolerance of ambiguity as a personality 
behaviour in syllogistic reasoning tasks. Paper presented to 
the Experimental Psychological Association.


Fraser, L.T. (1966a) Belief, incongruity, and syllogistic reasoning. Psychological Reports, 18, 982.


<table>
<thead>
<tr>
<th>APPENDICES I - VI</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I The logic of syllogisms.</td>
<td>2</td>
</tr>
<tr>
<td>II Example problems.</td>
<td>13</td>
</tr>
<tr>
<td>III Example passages.</td>
<td>14</td>
</tr>
<tr>
<td>IV Believability ratings of conclusions used in Experiments 1-3, 5 and 7.</td>
<td>20</td>
</tr>
<tr>
<td>V Dickstein's (1981) instructions on the interpretation of necessity.</td>
<td>23</td>
</tr>
<tr>
<td>VI On the conflict between logic and belief in syllogistic reasoning. (A paper based on Experiments 1 to 3).</td>
<td>26</td>
</tr>
</tbody>
</table>
APPENDIX I

THE LOGIC OF SYLLOGISMS

The categorical syllogism.

A syllogism is a deductive argument in which a conclusion is inferred from two premises. A categorical syllogism consists of three categorical propositions which contain three terms, each of which appears in two of the three constituent propositions. It is said to be in standard form when its premises and conclusion are all standard form categorical propositions and are arranged in a specified standard order. The following argument is a standard form categorical syllogism:

No A are B
All B are C

Therefore: Some C are not A

The three connected propositions of which this syllogism is composed, comply with the following structural conditions:
(a) Their terms, numerically six, must actually be three terms, each occurring twice.
(b) The subject of the conclusion must appear in one of the premises, usually the second.
(c) The predicate of the conclusion must appear in the other premise.
(d) The third term, known as the middle term, must occur in both
premises, but not in the conclusion. The syllogism can thus be seen as a connected whole; for the middle term joins together the two premises, and the two remaining (end) terms join the premises to the conclusion.

The two premises are distinguished by the technical terms; Major and Minor premise. The premise which contains the subject of the conclusion is called the Minor premise and its end term, the Minor term. Similarly, the premise which contains the predicate of the conclusion is called the Major premise, again its end term is called the Major term. The accepted order of propositions is Major premise, Minor premise followed by conclusion. The order of propositions does not, however, tell us anything about the validity of the syllogism. The validity depends only upon its form.

The form of a syllogism depends upon two things. Firstly, the types of categorical propositions contained within the syllogism; this specifies the mood of the syllogism. Secondly, the disposition of terms in the syllogism which specifies the figure of the syllogism. In order to determine the validity of any particular syllogism both the mood and the figure of that syllogism must be identified.

**Mood**

The mood of a syllogism is represented by three letters, each of which represents the type of proposition in the major premise, the minor premise and the conclusion. Categorical syllogisms can contain any combination of four possible propositional forms which are denoted by letters of the alphabet:
(A) Universal affirmative  All A are B
(E) Universal negative  No A are B
(I) Particular affirmative  Some A are B
(O) Particular negative  Some A are not B

For example, the mood of the following syllogism is stated as AOO, where the major premise is a universal affirmative proposition, the minor premise is a particular negative proposition and the conclusion is also a particular negative proposition:

All sprinters are athletes
Some gymnasts are not athletes

Therefore: Some gymnasts are not sprinters

Consider also an alternative AOO syllogism. This is identical in terms of mood. It differs, however, in figure.

All sprinters are athletes
Some athletes are not gymnasts

Therefore: Some gymnasts are not sprinters
Figure

Of the two types of syllogism given above, the first is valid and the second is invalid. It is thus demonstrated that the validity of any syllogism cannot be determined by its mood alone, but by a combination of its mood and figure. There are, in fact, four different figures which can be symbolised in the following way:

<table>
<thead>
<tr>
<th>FIGURE 1</th>
<th>FIGURE 2</th>
<th>FIGURE 3</th>
<th>FIGURE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>M - P</td>
<td>P - M</td>
<td>M - P</td>
<td>P - M</td>
</tr>
<tr>
<td>S - M</td>
<td>S - M</td>
<td>M - S</td>
<td>M - S</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>S - P</td>
<td>S - P</td>
<td>S - P</td>
<td>S - P</td>
</tr>
</tbody>
</table>

Where S denotes the subject of the conclusion, P denotes the predicate and M denotes the middle term which appears in and links both premises. If we return to our two example syllogisms, above, we see that the valid AOO syllogism is a figure two syllogism. The invalid AOO syllogism is a figure four syllogism. The two syllogisms may thus be coded as AOO-2 and AOO-4 respectively.

Since each of the propositions can be any one of four possible types (A, E, I or O) it follows that 4x4x4 = 64 moods can be constructed for each figure. When all four figures are taken into account, it is commonly accepted that there are 64x4 = 256 different sorts of syllogism. However, this assumes that the order of premises must remain fixed. If the premises are interchanged so that the subject of the conclusion appears in the first premise this doubles the number of possible syllogisms.
Hence, we may have 512 different types of syllogism. In logical terms, altering the order of the premises does not affect the validity of a syllogism. However, in psychological terms the effect of premise order should not be underestimated.

**Truth and validity**

In standard propositional logic, arguments are analysed according to their logical form. The truth or falsity of individual propositions is irrelevant. The validity of a syllogism depends solely upon the relation between the premises and the conclusion. Validity can be seen as a property of arguments which are groups of statements. Truth, however, is a property of individual statements, not arguments. There are certain circumstances in which a valid syllogism must have a true conclusion. For instance, if both premises are true, then the conclusion must also be true. However, it is not the case that a syllogism is proved valid by virtue of having a true conclusion. For example, the following syllogism has a true conclusion but is, however, invalid.

All pianists are performers
Some performers are musicians

Therefore: Some musicians are pianists
The analysis of validity

Logicians generally agree that the validity of standard form categorical syllogisms can be established by applying one or more of a specific set of rules.

Two of the rules refer specifically to the term 'distribution' to characterise the ways in which terms can occur in categorical propositions. The concept of distribution may be explained briefly. A term is distributed in a proposition when the proposition refers to all members of the class designated by that term. Otherwise the term is said to be undistributed in (or by) that proposition. The distribution of terms in the four standard categorical propositions can be summarised as follows:

<table>
<thead>
<tr>
<th></th>
<th>UNIVERSAL AFFIRMATIVE</th>
<th>UNIVERSAL NEGATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>distributed</td>
<td>distributed</td>
</tr>
<tr>
<td>Predicate</td>
<td>undistributed</td>
<td>distributed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PARTICULAR AFFIRMATIVE</th>
<th>PARTICULAR NEGATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>undistributed</td>
<td>undistributed</td>
</tr>
<tr>
<td>Predicate</td>
<td>undistributed</td>
<td>distributed</td>
</tr>
</tbody>
</table>
Rules of distribution

1. The middle term must be distributed in at least one premise. If the middle term is undistributed, then, in each of the premises, it may apply to different parts or members of the same whole and thus fail to bring the two end terms together. For example, if some men are hunters and some men are farmers, it does not necessarily follow that some hunters are farmers. It is also acceptable to say that the hunters are not the men who are farmers. Both conclusions are compatible with the premises, yet they are clearly contradictory.

2. A term that is distributed in the conclusion must be distributed in the corresponding premise. This rule simply states that a term cannot be used in a wider sense in the conclusion than in its corresponding premise.

Corollary
From rules 1 and 2 together it follows that there must be at least one more distributed term in the premise than in the conclusion.
Rules of quality and quantity

The distinction between affirmative and negative is called a distinction of Quality. The distinction between universal and particular is a distinction of Quantity.

3. At least one premise must be affirmative
A negative proposition separates its terms. Where both premises are negative, both end terms are separated from the middle term and hence no conclusion can be drawn concerning the two end terms.

4. A negative conclusion cannot follow from two affirmative premises.
When both premises are affirmative, both end terms are included in the middle term. Therefore, it cannot be concluded that one end term excludes the other.

5. If either premise is negative, the conclusion must be negative.
If a relation is affirmed between one end term, but denied between the other end term and the middle, then if any conclusion can be drawn, it must be one denying the relation between the two end terms.

6. One premise, at least, must be universal.

7. If one premise is particular, the conclusion must be
Rules 6 and 7 are corollaries of rules 1 to 5 and, therefore, can be proved from them.

**The psychological question**

Although the knowledge and application of the above rules is a recognised way of syllogistic analysis, it is not the application of these rules which interests psychologists as such. Psychologists seek to discover the means by which people perform syllogistic reasoning tasks without explicit knowledge of these rules.

A substantial proportion of theoretical investigation has focused on the possibility of errors due to different types of misinterpretation. There are two main possibilities here. People may misinterpret the logical meaning of quantifiers. They may also fail to interpret necessity in a strictly logical sense.

Let us examine how this could be so.

**The interpretation of quantifiers:**

**the relationship between two sets.**

In all there are five possible relations that may obtain between the two terms of a proposition. These relations may be expressed in Euler diagrams.
In order to analyse a proposition, each and every consistent set relation must be considered. In addition, the reasoner must appreciate that only two of the four propositions, given above, imply their converse. These are E and I propositions. Depending on how propositions are encoded and represented, it may not always be clear that a given proposition may not be converted. This may lead to errors.

An alternative source of error may lie in the misinterpretation of logical necessity. In standard propositional reasoning, a conclusion is said to follow from two premises if and only if it follows necessarily from every pair of set relations entailed by those premises. A conclusion which is possible, but not necessary cannot be termed a valid conclusion.
The categorical syllogism is used essentially by psychologists as a tool to help them answer a variety of questions. The interests of psychologists who study syllogistic reasoning have largely centred around the issue of rationality. The fundamental question here is whether errors (and correct responses) are due to logical or non-logical processes. Various theories have been put forward to account for performance on syllogistic reasoning tasks. They differ greatly in the degree of emphasis which is placed on logical processes.
APPENDIX II

EXAMPLE PROBLEMS

Example problems demonstrating the four content types and the four problem types used in Experiment 1. All problems are in the form of EIO-2 syllogisms. Believability is manipulated by interchanging the A and C terms within the conclusion. Validity is manipulated similarly.

VALID-BELIEVABLE
No police dogs are vicious
Some highly trained dogs are vicious

Some highly trained dogs are not police dogs

VALID-UNBELIEVABLE
No nutritional things are inexpensive
Some vitamin tablets are inexpensive

Some vitamin tablets are not nutritional

INVALID-BELIEVABLE
No addictive things are inexpensive
Some cigarettes are inexpensive

Some addictive things are not cigarettes

INVALID-UNBELIEVABLE
No millionaires are hard workers
Some rich people are hard workers

Some millionaires are not rich people

NOTE: Each of the above syllogisms was embedded in a prose passage.
The following passages are examples of the prose passages employed in Experiments 1 to 3. Each passage contains a pair of syllogistic premises. It is followed by the conclusion to be evaluated.

**PROBLEM TYPE: VALID-BELIEVABLE CONCLUSION (L+B+)***

The Health Education Council has recently expressed concern over the increase in addiction to various substances. It is believed that people begin with cigarettes and then progress to harder drugs. It has been suggested that the price of cigarettes could be increased to dissuade people from smoking in the first place. Some addictive things are inexpensive, and there is still growing demand for such things. No cigarettes are inexpensive, however, it is difficult to say what plays the major role in addiction, price may be only one of several factors.

Given that the above passage is true, does it follow that:

**SOME ADDICTIVE THINGS ARE NOT CIGARETTES**
PROBLEM TYPE: VALID-UNBELIEVABLE CONCLUSION (L+B-)  

Dogs are used extensively for the purpose of guarding property, guiding the blind and so on. Some police dogs are vicious, and many people believe that their temperament cannot be trusted. The police service use dogs a great deal in their work. No highly trained dogs are vicious and although fatal accidents are rare, there is still growing concern over their widespread use.

Given that the above passage is true, does it follow that:
SOME POLICE DOGS ARE NOT HIGHLY TRAINED

PROBLEM TYPE: INVALID-BELIEVABLE CONCLUSION (L-B+)

In a recent report to the World Health Organisation, an investigating committee suggested that the diet of many third world countries was nutritionally deficient. Although aid could be sent to such countries, no nutritional things are inexpensive. In the present economic climate cost is an important factor. Other types of aid could be sent instead and this would reduce the cost dramatically. Since some vitamin tablets are inexpensive, there must indeed be more viable alternatives.

Given that the above passage is true, does it follow that:
SOME NUTRITIONAL THINGS ARE NOT VITAMIN TABLETS
A group of sociologists have recently published a study on the relationship between wealth and hard work. Their results show that no millionaires are hard workers; they spend most of their time relaxing or being entertained. Some rich people are hard workers, in fact they are probably amongst the hardest workers in the country. Little is known about the standard of work associated with poorer people, but it is hoped that a subsequent study will concentrate specifically on these people.

Given that the above passage is true, does it follow that:
SOME MILLIONAIRES ARE NOT RICH

Since the turn of the century the number of young men joining the clergy has fallen dramatically. It has been suggested that this is due to the lure of scientific achievements, which tends to draw people away from religion. Some scientifically qualified people are religious, and such people are renowned for their tolerance of ideas. However, no scientifically qualified people are priests; indeed many people believe that the two views of life are irreconcilable. This view is stunning in its implications since, if it is taken to its extreme, it seems to suggest that the future is bleak for religion and sooner or later it may even disappear completely.
Given that the above passage is true, does it follow that:

SOME RELIGIOUS PEOPLE ARE NOT PRIESTS

PROBLEM TYPE: VALID-UNBELIEVABLE CONCLUSION (L+B-)

With increasing commercial interest in profits from North Sea Oil, various companies are looking for cheaper devices to aid them in their quest for richer gains. Skilled workers are employed at great cost to operate in underwater crews. Some crewmen are deep sea divers and most of them are familiar with the dangers of this type of work. Although safety precautions are taken, no crewmen are good swimmers, and the underwater currents are very strong. Companies are, however, eager to make the work more mechanised; thereby reducing the number of men needed and hence the cost. Divers argue, however, that replacing men with machines would create even greater risks for the remaining team members and, therefore, plan to resist their introduction at all costs.

Given that the above passage is true, does it follow that:

SOME DEEP SEA DIVERS ARE NOT GOOD SWIMMERS
PROBLEM TYPE: INVALID-BELIEVABLE CONCLUSION (L-B+)

For several years now NASA have been producing more sophisticated and extremely expensive aircraft designed for various kinds of space travel. Although thousands volunteer for training, less than one percent are chosen. No initial volunteers are healthy people, but even if they are already trained they still have to be put through their paces in gruelling machines and subjected to intense stress. According to detailed reports, some initial volunteers are astronauts, but this is not the major factor in selection. In spite of this rigorous testing, applications continue to increase. This is, however, a situation which NASA are eager to sustain, since such numbers ensure that future standards within the organisation will continue to be high and may even improve.

Given that the above passage is true, does it follow that:

SOME HEALTHY PEOPLE ARE NOT ASTRONAUTS
As a result of the rapidly increasing crime rate, the government have decided to build several more prisons to accommodate the rise in inmates. Critics claim, however, that the government are ignoring the social problems which may lie at the root of crime. A committee has, therefore, been set up to look into such problems. No members of the committee are judges, but people from a variety of backgrounds played a role in setting up the group. Some members of the committee are well educated, but this may be advantageous, as it may prevent their decisions from being biased against any one social class; as some critics have feared. Although the study is not yet finished, the findings so far are quite enlightening and may even change judicial policy.

Given that the above passage is true, does it follow that:

SOME JUDGES ARE NOT WELL EDUCATED PEOPLE
APPENDIX IV

BELIEVABILITY RATINGS OF CONCLUSIONS USED IN EXPERIMENTS 1 TO 3.

<table>
<thead>
<tr>
<th>Materials</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A (Experiments 1, 2 and 3)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T High trained dogs are not police dogs</td>
<td>6.44</td>
<td>0.89</td>
</tr>
<tr>
<td>F Some police dogs are not high trained</td>
<td>2.75</td>
<td>1.84</td>
</tr>
<tr>
<td>T Some nutritional things are not vitamin tablets</td>
<td>5.75</td>
<td>2.11</td>
</tr>
<tr>
<td>F Some vitamin tablets are not nutritional things</td>
<td>3.81</td>
<td>1.64</td>
</tr>
<tr>
<td>T Some addictive things are not cigarettes</td>
<td>6.25</td>
<td>1.88</td>
</tr>
<tr>
<td>F Some cigarettes are not addictive things</td>
<td>2.81</td>
<td>1.64</td>
</tr>
<tr>
<td>T Some rich people are not millionaires</td>
<td>5.94</td>
<td>1.57</td>
</tr>
<tr>
<td>F Some millionaires are not rich people</td>
<td>3.00</td>
<td>1.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B (Experiment 3)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T Some religious people are not priests</td>
<td>6.94</td>
<td>0.25</td>
</tr>
<tr>
<td>F Some priests are not religious people</td>
<td>1.69</td>
<td>1.14</td>
</tr>
<tr>
<td>T Some healthy people are not astronauts</td>
<td>6.94</td>
<td>0.25</td>
</tr>
<tr>
<td>F Some astronauts are not healthy people</td>
<td>3.75</td>
<td>2.18</td>
</tr>
<tr>
<td>T Some good swimmers are not deep sea divers</td>
<td>6.31</td>
<td>1.08</td>
</tr>
<tr>
<td>F Some deep sea divers are not good swimmers</td>
<td>2.75</td>
<td>2.18</td>
</tr>
<tr>
<td>T Some well educated people are not judges</td>
<td>6.44</td>
<td>1.55</td>
</tr>
<tr>
<td>F Some judges are not well educated people</td>
<td>3.31</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Note: All items were rated on a seven point scale from 1=completely unbelievable to 7=completely believable.
Materials A were rated by two groups of 16 subjects and Materials B by two separate groups of 16 subjects.
Each subject rated four statements, one from each context, of which two were "true" and two "false".
<table>
<thead>
<tr>
<th>Materials</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T Some musical instruments are not french horns</td>
<td>6.88</td>
<td>0.34</td>
</tr>
<tr>
<td>F Some french horns are not musical instruments</td>
<td>3.38</td>
<td>0.34</td>
</tr>
<tr>
<td>T Some books are not novels</td>
<td>6.88</td>
<td>0.34</td>
</tr>
<tr>
<td>F Some novels are not books</td>
<td>2.44</td>
<td>2.03</td>
</tr>
<tr>
<td>T Some sweets are not chocolates</td>
<td>6.81</td>
<td>0.40</td>
</tr>
<tr>
<td>F Some chocolates are not sweets</td>
<td>3.50</td>
<td>2.34</td>
</tr>
<tr>
<td>T Some wealthy people are not rich actors</td>
<td>6.94</td>
<td>0.25</td>
</tr>
<tr>
<td>F Some rich actors are not wealthy people</td>
<td>3.00</td>
<td>2.31</td>
</tr>
<tr>
<td>T Some antique furniture is not victorian furniture</td>
<td>6.94</td>
<td>0.25</td>
</tr>
<tr>
<td>F Some victorian furniture is not antique furniture</td>
<td>3.19</td>
<td>2.17</td>
</tr>
<tr>
<td>T Some serious accidents are not fatal accidents</td>
<td>6.13</td>
<td>1.31</td>
</tr>
<tr>
<td>F Some fatal accidents are not serious accidents</td>
<td>1.38</td>
<td>1.03</td>
</tr>
<tr>
<td>T Some animals are not dogs</td>
<td>6.56</td>
<td>1.50</td>
</tr>
<tr>
<td>F Some dogs are not animals</td>
<td>1.69</td>
<td>1.30</td>
</tr>
<tr>
<td>T Some vehicles are not cars</td>
<td>6.94</td>
<td>0.25</td>
</tr>
<tr>
<td>F Some cars are not vehicles</td>
<td>2.38</td>
<td>2.16</td>
</tr>
<tr>
<td>T No happy people are sad people</td>
<td>4.94</td>
<td>1.77</td>
</tr>
<tr>
<td>T No sad people are happy people</td>
<td>5.06</td>
<td>1.34</td>
</tr>
<tr>
<td>T No careless people are competent surgeons</td>
<td>5.06</td>
<td>1.77</td>
</tr>
<tr>
<td>T No competent surgeons are careless people</td>
<td>5.00</td>
<td>1.71</td>
</tr>
<tr>
<td>T No tory ministers are communists</td>
<td>6.31</td>
<td>1.20</td>
</tr>
<tr>
<td>T No communists are tory ministers</td>
<td>5.88</td>
<td>1.59</td>
</tr>
<tr>
<td>T No gymnasts are unfit people</td>
<td>5.88</td>
<td>1.41</td>
</tr>
<tr>
<td>T No unfit people are gymnasts</td>
<td>6.19</td>
<td>1.17</td>
</tr>
</tbody>
</table>

Note: All items were rated by 16 subjects. Each subject rated a statement from each type of content. Half of the statements were "true" and half were "false". A seven point rating scale was employed, as for Experiments 1 to 3.
### BELIEVABILITY RATINGS OF CONCLUSIONS USED IN EXPERIMENT 6

<table>
<thead>
<tr>
<th>Materials</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T Some mammals are not whales</td>
<td>2.17</td>
<td>1.92</td>
</tr>
<tr>
<td>F Some whales are not mammals</td>
<td>-2.39</td>
<td>1.54</td>
</tr>
<tr>
<td>T Some flowers are not daffodils</td>
<td>2.67</td>
<td>1.41</td>
</tr>
<tr>
<td>F Some daffodils are not flowers</td>
<td>-2.28</td>
<td>1.93</td>
</tr>
<tr>
<td>T Some insects are not bees</td>
<td>2.33</td>
<td>1.53</td>
</tr>
<tr>
<td>F Some bees are not insects</td>
<td>-2.06</td>
<td>1.86</td>
</tr>
<tr>
<td>T Some animals are not tigers</td>
<td>2.28</td>
<td>1.93</td>
</tr>
<tr>
<td>F Some tigers are not animals</td>
<td>-2.33</td>
<td>1.65</td>
</tr>
<tr>
<td>T Some vegetables are not parsnips</td>
<td>2.56</td>
<td>1.42</td>
</tr>
<tr>
<td>F Some parsnips are not vegetables</td>
<td>-2.89</td>
<td>0.47</td>
</tr>
<tr>
<td>T Some birds are not seagulls</td>
<td>3.00</td>
<td>0.00</td>
</tr>
<tr>
<td>F Some seagulls are not birds</td>
<td>-2.56</td>
<td>1.46</td>
</tr>
<tr>
<td>N Some members of the Bullman committee are not representatives of the Astrotex association</td>
<td>0.39</td>
<td>0.98</td>
</tr>
<tr>
<td>N Some representatives of the Astrotex association are not members of the Bullman committee</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>N Some collectors of Sparrowglaze pottery are not members of the Potpourri club</td>
<td>0.78</td>
<td>1.52</td>
</tr>
<tr>
<td>N Some members of the Potpourri club are not collectors of Sparrowglaze pottery</td>
<td>0.67</td>
<td>1.19</td>
</tr>
</tbody>
</table>

Note: All items were rated by 16 subjects. Each subject rated a statement from each type of content. Half of the statements were "true" and half were "false". A seven point rating scale was employed. In this case, the scale differed from that of Experiments 1 to 3, in that -3=completely unbelievable and +3=completely believable.
APPENDIX V

DICKSTEIN'S (1981) INSTRUCTIONS ON THE INTERPRETATION OF NECESSITY

Syllogistic Reasoning

On the next 16 pages you will find 64 syllogisms to solve. A syllogism is a pair of statements providing you with information about the relationships between three classes. These three classes are always labelled S, M, and P in the present study. One statement in the syllogism will always provide you with information about the relation between M and P. The other statement in the syllogism will always provide you with information about the relation between S and M.

On the basis of the information provided by the first two statements, you are asked to draw a conclusion regarding the relationship between S and P. Each time there are five possible alternatives and your task is to circle the correct alternative.

The five alternatives are:

1. All S are P
2. Some S are P
3. No S are P
4. Some S are not P
5. No valid conclusion
In every instance there is only one correct answer. It is essential for this research that you answer every item. Please make sure you do not omit any of the syllogisms.

In solving syllogisms, it is important to keep certain rules in mind. These rules specify the meaning of the terms and when conclusions may be drawn.

The first rule is that in order to draw one of the first four conclusions the information must allow that conclusion with certainty. The fact that a relationship is possible but not compelled by the information in the syllogism does not justify any of the first four conclusions. If there are possible relationships between S and P but no necessary relationship compelled by the syllogism, then the correct answer is alternative five - No valid conclusion.

The second rule is that the word "some" in a syllogism means "at least some" and does not mean "only some". Thus, when you endorse alternative two, you are concluding that at least some S are P but are not implying that some S are not P. Similarly, if you endorse alternative four, you are concluding that some S are not P but are not implying that some S are P.

The third rule is that when two conclusions are justified by the syllogism you must always draw the strongest conclusion. Thus, when a syllogism allows you to draw the conclusion that all S are P, then it is incorrect to conclude that some S are P even though this will always be true. Similarly, if a syllogism allows you to draw the conclusion that no S are P, then it is incorrect to conclude that some S are not P even though this will
always be true.

By keeping in mind this information, it will be possible for you to avoid common errors in syllogistic reasoning. Work carefully and try to reason each syllogism through before selecting a conclusion.

Thank you for your participation in this research.
APPENDIX VI

PAPER ENTITLED

ON THE CONFLICT BETWEEN LOGIC AND BELIEF IN SYLLOGISTIC REASONING.
On the conflict between logic and belief in syllogistic reasoning

J. St. B. T. EVANS, JULIE L. BARSTON, and PAUL POLLARD
Plymouth Polytechnic, Plymouth PL4 8AA, England

Three experiments are reported that investigate the weighting attached to logic and belief in syllogistic reasoning. Substantial belief biases were observed despite controls for possible conversions of the premises. Equally substantial effects of logic were observed despite controls for two possible response biases. A consistent interaction between belief and logic was also recorded; belief bias was more marked on invalid than on valid syllogisms. In all experiments, verbal protocols were recorded and analyzed. These protocols are interpreted in some cases as providing rationalizations for prejudiced decisions and, in other cases, as reflecting a genuine process of premise to conclusion reasoning. In the latter cases, belief bias was minimal but still present. Similarly, even subjects who focus primarily on the conclusion are influenced to an extent by the logic. Thus a conflict between logic and belief is observed throughout, but at several levels of extent.

An important debate in cognitive psychology surrounds the notion of rationality with respect to human inference (see Cohen, 1981, and associated commentaries). Recent reviews by Evans (1982) and Nisbett and Ross (1980) have stressed the role of apparently irrational processes in the study of inductive and deductive inference, respectively (see also Pollard, 1982). However, theories favoring a rationalist interpretation of inferential behavior still hold a dominant position in the recent literature (see, for example, the collections of papers edited by Falmagne, 1975, and Revlin & Mayer, 1978). In the case of deductive reasoning, much of the argument centers on the use by subjects of a system of logic, whether of the philosopher’s variety (cf. Henle, 1962) or of an alternative "natural" type (e.g., Braine, 1978).

The nonlogical or antirational position is sometimes misinterpreted as denigrating man’s proven intelligence. What is in fact suggested is an alternative conception of that intelligence. The “rationalist” position entails the supposition that the reasoner proceeds by forming an abstract representation of problem information and applying a general set of inferential rules to its logical structure, regardless of its content. This notion is clearly embodied, for example, in Piaget’s theory of formal operations (cf. Inhelder & Piaget, 1958). The alternative argument stressed here is that specific features of problem content, and their semantic associations, constitute the dominant influence on thought (see Evans, 1982, for extended discussion).

In this paper, we will focus on the alleged “belief-bias” effect in reasoning. The claim is that when presented with deductive arguments to evaluate, subjects will make judgments upon a priori beliefs rather than on the basis of logical argument. Specifically, they will tend to endorse arguments whose conclusions they believe and reject arguments whose conclusions they disbelieve, irrespective of their actual validity. A number of authors have claimed evidence of such an effect in syllogistic reasoning (e.g., Feather, 1964; Gorden, 1953; Henle & Michael, 1956; Janis & Frick, 1943; Janis & Terwuliger, 1962; Kaufman & Goldstein, 1967; Lefford, 1946; Morgan & Morton, 1944; Wilkins, 1928; Wilson, 1965; Winthrop, 1946).

Most of these studies are, however, open to criticism on a variety of grounds (cf. Evans, 1982). Some find only weak effects, not backed by tests of statistical significance; others use poorly worded problems, fail to control for other factors that influence reasoning, or risk carryover effects by having subjects rate the believability and validity of arguments in the same session. Revlin and Leirer (1978) and Revlin, Leirer, Yopp, and Yopp (1980) have raised other problems, such as a failure to control for the pragmatic convertibility of the logical premises, that may affect the logical status of the problem representation. Revlin et al. argue that a rational reasoner, as proposed by the model of Revlin (1975a, 1975b), should not manifest belief biases, and they rightly contend that the aging literature on the subject should be opened up to investigation with improved methodologies. While finding some evidence of belief-bias effects, Revlin et al. argue that these are relatively weak in comparison with the logical tendencies observed, when due allowance is made for personalized representations of the premises. The present study questions the accuracy of this conclusion.

First, we must briefly review the claims of the Revlin...
studies employed primarily valid syllogisms. There is no way to present only one conclusion for evaluation, as in the experiments to be reported here. Hence, the reasoner makes his or her judgments based upon "too much" information, and not only on the specific content of the problem. Revlin suggests that this is manufactured in terms of the kinds of immediate inferences that the reasoners are willing to make when presented with a proposition. For example, it is claimed that when shown the abstract relations "All A are B," reasoners often infer that "All B are A." However, due to pragmatic implications, conversion may be blocked in some concrete relations (e.g., "All dogs are animals"). It is this kind of use of personal knowledge that the conversion model claims will affect the validity judgments on categorical syllogisms. It is proposed that subjects will be correct in their judgments in either of two conditions: (1) when the premises of the problem are converted, but the logical conclusion is fortuitously the same in the converted and the original forms of the problem, and (2) when the subject's knowledge of the world blocks illicit conversion. Revlin et al. (1980; see also Revlin & Leirer, 1978) report evidence to support these hypotheses. As mentioned earlier, they also find significant, although they claim "weak," evidence of belief bias when conversion of the premises is fully controlled. However, there are several methodological problems identified by Pollard (1979) that may have led Revlin et al. to underestimate the extent of the belief-bias effect. For example, Revlin and Leirer (1978) claim that belief and logic conflict in the following problem: "No U.S. governors are members of the Harem Club. Some Arabian sheiks are members of the Harem Club. Therefore: (a) All Arabian sheiks are U.S. governors. (b) No Arabian sheiks are U.S. governors. (c) Some Arabian sheiks are U.S. governors. (d) Some Arabian sheiks are not U.S. governors. (e) None of the above is proved."

The "believable" answer is claimed to be Answer b, but Answer d is also empirically true. Since Answer d is the logically correct answer as well, Revlin and Leirer's (1978) claim that subjects are overriding personal beliefs in choosing it is doubtful. This problem arises as a result of the multiple-choice technique, and it can be avoided by presenting only one conclusion for evaluation, as in the experiments to be reported here. A second problem is that the Revlin (1975a, 1975b) studies employed primarily valid syllogisms. There is evidence to suggest that belief-bias effects may be more marked on indeterminate syllogisms, that is, on those whose conclusions do not follow logically (Kaufman & Goldstein, 1967).

A third problem is that Revlin and Leirer (1978) did not control for the effects of "atmosphere," an alleged bias to choose syllogistic conclusions that share syntactic features with the premises, regardless of logical validity. Although existence of this effect, proposed by Woodworth and Sells (1935), is regarded primarily as an artifact of conversion by Revlin (1975a, 1975b), there is much evidence compatible with the hypothesis (see Evans, 1982). Since Gorden (1953) has specifically suggested that belief bias is weaker than atmosphere, it is advisable to control for its possible effects. In fact, the logically correct answer to Revlin et al.'s (1980) valid syllogisms was also the conclusion favored by atmosphere, which may have led Revlin et al. to overestimate subjects' logical abilities.

Experiment 1 was designed to test the relative weighting given to logic and belief in syllogistic reasoning, controlling for conversion of premises as Revlin (1975a, 1975b) requires, but also improving upon his methodology in the three respects outlined above. In addition, subjects were asked to provide retrospective verbal justifications of their decisions. On a different reasoning task, Wason and Evans (1975) claimed that such reports indicated no evidence of insight into the origin of responses apparently induced by a nonlogical bias and produced logical sounding post hoc rationalizations. Further investigation of this phenomenon of interest with respect to the recent debate about the interpretation of verbal protocols (see, for example, Ericsson & Simon, 1980; Nisbett & Wilson, 1977).

EXPERIMENT I

The type of problems used in Experiment I was categorical syllogisms. A syllogism is a deductive argument consisting of two premises and a conclusion. The two premises make statements about the relations between three terms: a major term (P), a minor term (S), and a middle term (M).

The figure of a syllogism indicates the position of the middle term in the premises. There are four possible figures, shown in Figures 1-4.

Syllogisms are composed of a combination of four basic types of statement: (1) The universal affirmative statement: All X are Y (symbolized by "A"). (2) The particular affirmative statement: Some X are Y (symbolized by "I"). (3) The universal negative statement: No X are Y (symbolized by "E"). (4) The particular negative statement: Some X are not Y (symbolized by "O"). The types of statement that occur in any particular syllogism specify its mood.

The form of a syllogism may be completely described, therefore, by stating its mood and figure. Thus a (valid) EIO-2 syllogism is of the following form: "No A are B.
Some C are B. Therefore, some C are not A.” This was one of the types of syllogism used in Experiment 1.

Both E and I statements are “legally” convertible, which means that the terms of the statement may be reversed without altering its meaning in logic. Since the two premises of the syllogisms used in this experiment are of this type, even if subjects do in fact convert premises, as Revlin and his associates suggest, this could not in itself be the cause of logical errors. The conversion model, therefore, predicts that reasoning on this task will be logical, regardless of the type of material used, and there should be no belief bias.

In the Introduction, a reference was made to atmosphere bias. This is a nonlogical bias first investigated by Woodworth and Sells (1935; see also Begg & Denny, 1969). Briefly, this effect is due to the type of quantifiers used in the two premises, which combine to create an “atmosphere,” which predisposes subjects to accept a conclusion containing specific quantifiers. Since the conclusion to EI premises favored by atmosphere bias is an O conclusion, all conclusions are favored by the bias and it cannot confound any comparisons between problems.

### Method

Materials. Half of the syllogisms presented were valid, and half were invalid (i.e., their conclusions did not follow logically from the premises). The following two syllogisms were used throughout: “No A are B. Some C are B. Therefore, some C are not A” (valid). “No A are B. Some C are B. Therefore, some A are not C” (invalid).

Note that invalid conclusions were thus of the form C-A and invalid, A-C. Both syllogisms are in Figure 2, but the latter reverses the traditional premise order, so that the mood remains the same for both valid and invalid syllogisms.

The materials were chosen so that the conclusions of the syllogisms would appear “true” when the premises were presented in one order, but “false” when the order of terms was reversed. The experimenters’ intuitions were checked by having a group of 32 subjects, who did not participate in the experiment, rate the conclusions for believability (see Table I). It will be seen that the differences in ratings between “true” and “false” sentences are very marked.

Half of the valid conclusions presented to subjects were believable, and half were unbelievable. For example, the following syllogism, which is valid, has a believable conclusion: “No cigarettes are inexpensive. Some addictive things are inexpensive. Therefore, some addictive things are not cigarettes.” This valid syllogism, on the other hand, has an unbelievable conclusion: “No addictive things are inexpensive. Some cigarettes are inexpensive. Therefore, some cigarettes are not addictive.”

For invalid syllogisms, as for valid, half were believable, and half were not: “No addictive things are inexpensive. Some cigarettes are inexpensive. Therefore, some cigarettes are not addictive.”

### Table I

<table>
<thead>
<tr>
<th>Materials</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Experiments 1, 2, and 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Some highly trained dogs are not police dogs</td>
<td>6.44</td>
<td>.89</td>
</tr>
<tr>
<td>F Some police dogs are not highly trained</td>
<td>2.75</td>
<td>1.84</td>
</tr>
<tr>
<td>T Some nutritional things are not vitamin tablets</td>
<td>5.75</td>
<td>2.11</td>
</tr>
<tr>
<td>F Some vitamin tablets are not nutritional things</td>
<td>3.81</td>
<td>1.64</td>
</tr>
<tr>
<td>T Some addictive things are not cigarettes</td>
<td>6.25</td>
<td>1.88</td>
</tr>
<tr>
<td>F Some cigarettes are not addictive things</td>
<td>2.81</td>
<td>1.64</td>
</tr>
<tr>
<td>T Some rich people are not millionaires</td>
<td>5.94</td>
<td>1.57</td>
</tr>
<tr>
<td>F Some millionaires are not rich people</td>
<td>3.00</td>
<td>1.90</td>
</tr>
<tr>
<td>B (Experiment 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Some religious people are not priests</td>
<td>6.94</td>
<td>.25</td>
</tr>
<tr>
<td>F Some priests are not religious people</td>
<td>1.69</td>
<td>1.14</td>
</tr>
<tr>
<td>T Some healthy people are not astronauts</td>
<td>6.94</td>
<td>.25</td>
</tr>
<tr>
<td>F Some astronauts are not healthy people</td>
<td>3.75</td>
<td>2.18</td>
</tr>
<tr>
<td>T Some good swimmers are not deep sea divers</td>
<td>6.31</td>
<td>1.08</td>
</tr>
<tr>
<td>F Some deep sea divers are not good swimmers</td>
<td>2.75</td>
<td>2.18</td>
</tr>
<tr>
<td>T Some well educated people are not judges</td>
<td>6.44</td>
<td>1.35</td>
</tr>
<tr>
<td>F Some judges are not well educated people</td>
<td>3.31</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Note—All items were rated on a 7-point scale from 1 = completely unbelievable to 7 = completely believable. Materials A were rated by two groups of 16 subjects and Materials B by two separate groups of 16 subjects. Each subject rated four statements, one from each context, of which two were “true” and two “false.”
not cigarettes" (invalid, believable conclusion). "No cigarettes are inexpensive. Some addictive things are inexpensive. Therefore, some cigarettes are not addictive" (invalid, unbelievable conclusion).

There were thus four types of problem: valid conclusion, believable or unbelievable, and invalid conclusion, believable or unbelievable. Of course, the four problems actually given to each subject used different problem contents.

To reduce the artificiality of the task, problems were presented in the form of prose passages that were approximately 80 words in length. Four different types of passage content were used, each taking the form of a current affairs article. The four topics were: (1) public response to the behavior of police dogs, (2) the provision of aid for third-world countries, (3) attempts to reduce the number of people smoking cigarettes, and (4) the relationship between wealth and hard work. The following is an example of Passage Type I: "Dogs are used extensively for the purpose of guarding property, guiding the blind and so on. No highly trained dogs are vicious. However, many people believe that their temperament cannot be trusted. The police service uses dogs a great deal in their work. Some police dogs are vicious and although fatal accidents are rare, there is still growing concern over their widespread use." "If the above passage is true, does it follow that: Some highly trained dogs are not police dogs?" (This conclusion is invalid, but believable.)

Design. Each subject received each of the four types of prose passages and problem types, solving four problems in all. Combination of problem type and passage type was balanced in a Latin square design, and presentation order was randomized.

Subjects. Twenty-four undergraduates at Plymouth Polytechnic acted as paid volunteers. They had no previous experience of syllogistic reasoning tasks and were tested individually.

Procedure. Task and instructions. The instructions and problems for each subject were presented on typed cards. All problems were presented individually, and each problem card remained in front of subjects for reference when decisions were cast. The instructions began as follows: "This is an experiment to test people's reasoning ability. You will be given four problems. In each case, you will be given a prose passage to read and asked if a certain conclusion may be logically deduced from it. You should answer this question on the assumption that all the information given in the passage is, in fact, true. If you judge that the conclusion necessarily follows from the statements in the passage, you should answer 'yes,' otherwise 'no.'

"Please take your time and be sure that you have the right answer before stating it. When you have decided, I will then ask you to explain why you believe the conclusion to be valid or invalid as the case may be. Any questions?"

Subjects' protocols were recorded on a tape recorder for later analysis.

Protocol scoring. Each protocol was scored on a yes/no basis on two criteria: (1) presence or absence of a reference to both the logically relevant premises, and (2) presence or absence of references to irrelevant information, either within the passage or extraneous.

Results and Discussion

The percentage frequencies of subjects accepting the conclusion (i.e., deeming the argument to be valid) are shown for each type of problem in Table 2. As predicted, there was a substantial effect of "belief bias" (i.e., a tendency over all problems to accept more believable than unbelievable conclusions; p < .01, one-tailed sign test). There was also a tendency to accept more valid than invalid arguments overall (p < .02, one-tailed) and a significant Belief by Validity interaction (p < .05). The nature of this interaction accords with the findings of Kaufman and Goldstein (1967) that belief bias is more marked for invalid than for valid syllogisms.

These results cannot be reconciled with the essentially rationalist approach of Reuvlin and Leirer (1978) and Reuvlin et al. (1980). Their claim that previous evidence of belief bias could be an artifact of uncontrolled premise conversion cannot be applied to the present experiment, in which only E and I premises were involved. Their further suggestion, on the basis of their own data, that belief biases are weak in comparison with rational processes is also inconsistent with our results. The suggestion in the introduction that their methodology led to an underestimate of the true extent of belief bias is strongly confirmed in our data. They found subjects to be correct when logic accorded with belief on 83% of occasions, and when logic conflicted with belief, on more than 67% of occasions. The corresponding percentages in Experiment I were 92% and 27%.

The present results do accord well with the Evans (1982) two-factor theory, which claims that reasoning responses reflect a competition between logical and nonlogical tendencies. This theory has previously been applied mostly to conditional reasoning problems (e.g., Evans, 1977a, 1977b). Related to this is the dual process theory of Wason and Evans (1975), who claim that the verbalizations observed on their reasoning problems reflect primarily a type of thought different from that determining the reasoning response. Specifically, they found that subjects tended to rationalize responses attributed to nonlogical biases. Similar trends should be found in the protocols collected in Experiment 1.

However, Ericsson and Simon (1980) have argued persuasively that protocols may reveal the locus of the subjects' attention, or the information heeded by the problem solver. It could be that the Wason and Evans' (1975) rationalizations were due to their asking the subjects to justify the responses given. The important issue in the present study is that of whether subjects base their reasoning on the logical premises or on extraneous beliefs. The protocols were consequently scored separately for the presence or absence of references to both, and the results are shown in Table 3.

Of particular interest are the two conditions in which logic and belief conflict. If subjects are rationalizing, then we might expect that their protocol ratings would interact with their response to the problem. That is, subjects who give the logical response should make more references to the premises and those favoring belief

| Table 2 Percentage Frequency of Subjects Accepting Conclusions in Experiment I (n = 24) |
|---------------------------------|------------------|------------------|
|                                | Believable | Unbelievable |
| Valid                           | 92         | 46             |
| Invalid                         | 92         | 8              |
should make more references to irrelevant information. The same prediction would also be made if the protocols were assumed to reflect the actual information on which the subjects' reasoning was based (an attempt will be made to distinguish these possibilities in the general discussion). Only the valid-unbelievable condition produced a sufficiently even split of "yes" and "no" responses to permit test of this hypothesis. The predicted interaction was present and significant for references to irrelevant information (p = .026, one-tailed Fisher exact probability test), but not for references to the logical premises. It is also relevant to note that in the invalid-believable condition, subjects who accepted the conclusion (in accordance with beliefs) showed the highest ratio of references to irrelevant information and the lowest ratio of references to the logical premises observed in the whole experiment.

**EXPERIMENT 2**

Experiment 2 was designed to replicate and extend Experiment 1. There are several difficulties of interpretation of Experiment 1 that Experiment 2 was intended to resolve. First, the belief bias observed could be due to the embedding of the logical premises in a prose passage. This could decrease the subjects' attention to the logically critical premises. Hence, a group was included that was given the premises only. Second, it is possible that the instruction to give verbal justifications, especially with a within-subjects design, could affect responses. Hence, another group was added with no instructions to verbalize. Another group was used to investigate further the causes of the trends in the protocol analysis. This group provided "thinking-aloud" protocols, which Ericsson and Simon (1980) argue are more likely to give an accurate picture of the information heeded than is the retrospective method used in Experiment 1.

Finally, the problem structures were modified to take account of the "figural bias" discussed by Johnson-Laird and Steedman (1978). They show that the order in which terms are arranged in the syllogisms can exert an influence on the choice of conclusion, irrespective of logical validity. In Experiment 1, all premise pairs were of the form A-B, C-B, but valid conclusions were always of the form A-C and invalid conclusions of the form A-C. According to Johnson-Laird and Steedman's findings (but not their model), there may be a bias to prefer C-A conclusions with these premise types. This may have led us to overestimate subjects' logical ability in Experiment 1. In Experiment 2, both valid and invalid problems were associated with both A-C and C-A conclusions.

**Method**

Materials. The syllogisms and prose passages used were the same as those of Experiment 1, except for the modification that permitted both valid and invalid syllogisms to have both A-C and C-A conclusions. This was produced by interchanging the quantifiers of each of the original problem premises and reversing their conclusion, for example: "No A are B. Some C are B. Therefore, some C are not A." (original valid syllogism). "Some A are B. No C are B. Therefore, some A are not C." (valid control syllogism).

Design. As in Experiment 1, all subjects received four problems consisting of all four problem types combined with all four passage contents. In this experiment, four subject groups were used: Group 1 received prose passages and was required to verbalize the explanation for the decision retrospectively (as in Experiment 1). Group 2 was required to verbalize in the same manner as Group 1, but subjects received only the logical premises and not the full prose passage. Group 3 received prose passages and was required to verbalize concurrently (i.e., to think aloud while attempting to solve the problem). Group 4 received prose passages but was not required to verbalize at all.

Each of these four groups was then subdivided into two further groups, one of which received only A-C conclusions for both valid and invalid problems, and the other of which received C-A conclusions only.

Subjects. Sixty-four undergraduates at Plymouth Polytechnic acted as paid volunteers. They had no previous experience of this task and were tested individually.

Procedure. The instructions and problems were presented in the same manner as Experiment 1.

The instructions were as follows: Group 1—instructions as for Experiment 1. Group 2—instructions as for Experiment 1, except that any reference to the prose passage was omitted. Group 3—instructions as for first paragraph of Experiment 1, continued as follows, "Whilst you are trying to solve each problem 1 would like you to try to 'think aloud' as much as you can. Please don't let this distract you from the task in hand, which is to obtain the correct solution to the problem. If at any time during the task, I do not think that you are speaking enough, I will simply prompt you to speak a little more. Please take your time and be sure that you have the right answer before stating it. Any questions?" Group 4—instructions as for Experiment 1, except that any request to verbalize was omitted.

As in Experiment 1, subjects' protocols were recorded on a tape recorder for later analysis. Protocols were scored using the same procedure as in Experiment 1.

**Results and Discussion**

The percentage frequency of subjects accepting
arguments as valid in Experiment 2 is shown in Table 4. As is apparent from Table 4a, the order of terms in the conclusion had no significant effect on responses, and further analyses were collapsed over this factor. Binomial tests on the combined data of the subjects in all groups (n = 64) yielded highly significant evidence of the three effects found in Experiment 1. That is, more believable than unbelievable conclusions were accepted (p < .001), more valid than invalid conclusions were accepted (p < .001), and the two factors interacted (p < .01). The interaction reflects the fact that the belief-bias effect is more marked for invalid than for valid syllogisms. The interpretation of this interaction will be deferred to the general discussion. Overall, subjects were correct 87% of the time when logic accorded with belief and 48% of the time when it did not.

The same trends were manifested in each of the four groups (see Table 4b). In order to test whether response patterns were affected by group, a set of four 2 by 4 chi-square tests were carried out to compare yes/no frequencies across the four groups for each problem type. None of these analyses yielded a significant result.

The results of the protocol analyses are shown in Table 5, for the three groups from whom protocols were collected. Inspection of Table 5 suggests that the distribution of classification frequencies is, in fact, very similar for all three groups. This was confirmed statistically by rank ordering the 16 cell frequencies for each group and assessing the similarity of the rank orderings by Kendall's coefficient of concordance (cf. Siegel, 1956). There was high and significant concordance for both mention of the premises (W = .792, p < .001) and mention of irrelevant information (W = .871, p < .001). Consequently, further analyses were performed on the combined data of all three groups.

These analyses revealed highly significant interactions between the answer given to the problem and the protocol classifications for the valid-unbelievable problems. Subjects accepting the valid conclusion against its believability made more references to the logical premises ($\chi^2 = 24.61, p < .001$) and fewer references to irrelevant information ($\chi^2 = 15.11, p < .001$).

The other problem for which interactions were observed was the invalid-unbelievable type. Only six subjects went against both logic and belief to accept such arguments as valid. However, all six referred to both logical premises, and none referred to irrelevant information. Fisher exact probability tests revealed a significant interaction with the majority "no" responders in each case ($p < .005$ and $p < .05$, respectively). The simplest interpretation of these findings is that these subjects ignored beliefs and reasoned from the premises, but they did so with faulty logic.

### Table 4

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusion</td>
<td>B U</td>
<td>B U</td>
</tr>
<tr>
<td>C-A</td>
<td>84 68</td>
<td>59 13</td>
</tr>
<tr>
<td>A-C</td>
<td>88 56</td>
<td>72 13</td>
</tr>
<tr>
<td>Mean</td>
<td>86 62</td>
<td>66 13</td>
</tr>
</tbody>
</table>

Note - 1 = prose passage, retrospective verbalization; 2 = premises only, retrospective verbalization; 3 = prose passage, concurrent verbalization; 4 = prose passage, no verbalization. B = believable; U = unbelievable.

### Table 5

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Both Premises</th>
<th>Irrelevant Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
<td>Group 2</td>
</tr>
<tr>
<td></td>
<td>Y Y N N</td>
<td>Y Y N N</td>
</tr>
<tr>
<td>Valid-Believable</td>
<td>M 10 0 6 1 1</td>
<td>21 2 3 1 8 0 6 1 17 2</td>
</tr>
</tbody>
</table>

Note - M = mentioned; NM = not mentioned; Y = yes response; N = no response. Correct responses are italicized.
The interpretation of the protocol data will be taken up in the general discussion. There is one further problem concerning the interpretation of the decision frequencies that Experiment 3 was designed to deal with.

EXPERIMENT 3

In both Experiments 1 and 2, subjects accepted significantly more valid than invalid conclusions. This suggests that people have some ability to reason and overcome belief biases, at least for unbelievable conclusions. It is, however, possible that this apparent logicality is an artifact of a response bias different from that controlled in Experiment 2. In all the syllogisms used so far, the quantifier "some" always modified the same term in the premises as in the conclusion for valid problems (e.g., "No A are B. Some C are B. Therefore, some C are not A."). But for invalid problems, this was never the case (e.g., "No A are B. Some C are B. Therefore, some A are not C.").

It is therefore possible that some form of feature-matching bias is responsible for the main effect of validity. This problem can be overcome if syllogisms in Figure 3 rather than Figure 2 are employed. For example, the following is logically equivalent to the former of the two problems above: "No B are A. Some B are C. Therefore, some C are not A."

Since in all Figure 3 syllogisms the two terms used in the conclusion (A and C) appear in the premises, the possible response bias described could not operate. Experiment 3, therefore, compared subjects' performance on Figure 2 and Figure 3 syllogisms. Prose passages were employed, and thinking-aloud protocols were recorded.

Although the instructions presented in Experiments 1 and 2 clearly indicated that subjects' inferences should be based on logical necessity, this is an unusual requirement for subjects to follow. One interpretation of the belief-bias effect is that subjects "fail to accept the logical task" (Henle, 1962). In order to counter this possibility, the instructions of Experiment 3 were reworded to increase emphasis on the concept of logical necessity.

Method

Design. All subjects received eight problems to solve, the four types used previously in both Figure 2 and 3 syllogisms. Four further scenarios were constructed to add to the four used in Experiments 1 and 2, and each subject received each of the eight scenarios, randomly matched to the eight types of problems. The conclusion ratings for these additional problem contents are shown in Table 1. Presentation order was also randomized. All subjects were instructed to "think aloud" while solving the problems.

Subjects. Thirty-two 1st-year psychology students of Plymouth Polytechnic participated in partial fulfillment of course credit requirements. All were tested individually.

Procedure. The relevant section of the modified instructions follows: "Your task is to decide whether or not a given conclusion follows logically from the information given, and this information only. You must assume that all the statements within the passage are true; this is very important. If, and only if, you judge that the given conclusion logically follows from the statements given in the passage you should answer 'yes,' otherwise 'no.'"

The final sentence of this extract was repeated at the very end of the instructions. The procedure was otherwise similar to that of Experiments 1 and 2, with "thinking-aloud" instructions. Protocols were tape-recorded and subsequently transcribed and analyzed in a manner similar to that of the previous experiments.

Results

The frequencies of responses to the problems are shown in Table 6. It is apparent that there is no difference in performance between the Figure 2 and Figure 3 problems, thus eliminating the response-bias explanation of the validity effect. On the combined data, there were highly significant preferences to accept conclusions that were believable rather than unbelievable.

Table 6

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B U</td>
<td>B U</td>
</tr>
<tr>
<td>Figure 2</td>
<td>91</td>
<td>53</td>
</tr>
<tr>
<td>Figure 3</td>
<td>91</td>
<td>53</td>
</tr>
<tr>
<td>Combined</td>
<td>91</td>
<td>53</td>
</tr>
</tbody>
</table>

Note—B = believable; U = unbelievable.

Table 7

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Both Premises</th>
<th>Irrelevant Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Figure 2 Y N</td>
<td>Figure 3 Y N</td>
</tr>
<tr>
<td>Valid-Believable MM</td>
<td>5 2 11 0</td>
<td>16 1 17 1</td>
</tr>
<tr>
<td>NM</td>
<td>24 2 18 3</td>
<td>13 2 12 2</td>
</tr>
<tr>
<td>Valid-Unbelievable</td>
<td>5 8 7 11</td>
<td>9 8 9 6</td>
</tr>
<tr>
<td>NM</td>
<td>7 6 3 4</td>
<td>12 5 14 1</td>
</tr>
<tr>
<td>Invalid-Believable</td>
<td>15 4 18 7</td>
<td>10 5 7 6</td>
</tr>
<tr>
<td>NM</td>
<td>1 8 1 7</td>
<td>1 20 2 16</td>
</tr>
<tr>
<td>Invalid-Unbelievable</td>
<td>0 22 2 22</td>
<td>0 11 1 12</td>
</tr>
</tbody>
</table>

Note—M = mentioned; NM = not mentioned; Y = yes response; N = no response. Correct responses are italicized.
(p < .001, binomial test) and those that were valid rather than invalid (p < .001, binomial test). The interaction, although in the same direction as observed previously, fell just short of significance (p = .067, one-tailed binomial test). Subjects were correct 97% of the time when belief agreed with logic and 43% of the time when belief conflicted with logic.

The protocol analyses are summarized in Table 7. Previous experiments showed an interaction between the classification frequencies and type of response for the valid-unbelievable condition. A similar trend was observed in Experiment 3 on the references to premises criterion, although it fell short of significance for both figures. No interaction trend was apparent on the reference to irrelevant information criterion. No other conditions produced significant interactions.

GENERAL DISCUSSION

Over the three experiments, consistently large and significant effects of belief bias have been observed, despite controls for conversion of premises (cf. Revlin et al., 1980). Similarly, large and consistent effects of logical validity have been observed despite the controls introduced to test response-bias explanations in Experiments 2 and 3. There is also a consistent trend for the two factors to interact, such that the belief-bias effect is more marked for invalid than for valid problems. The strong instructional emphasis on logical necessity in Experiment 3 renders implausible any suggestion that the belief-bias effect reflects uncertainty on the subjects’ part of what they were required to do. If they are “failing to accept the logical task,” it is because they are unable to do so.

The Belief by Logic interaction arises because subjects respond differently to the two conditions in which logic and belief conflict. When the problem is invalid but believable, subjects generally accept the conclusion. Response rates are intermediate, however, when the syllogisms are valid but have unbelievable conclusions. This condition is especially interesting also with reference to protocol analysis, since subjects conforming to logic tend to refer to the premises, whereas those conforming to beliefs tend to refer instead to irrelevant information.

We must now ask what process of reasoning could account for these findings. There are a number of published models of syllogistic reasoning (e.g., Dickstein, 1978a, 1978b; Erickson, 1974; Guyote & Sternberg, 1981; Johnson-Laird & Steedman, 1978; Revlin, 1975a, 1975b). These models differ considerably in the details of their psychological descriptions, but in one respect they all agree. All the models suppose that the subject starts by forming a representation of the premises and then generates a conclusion, or set of possible conclusions, by a more or less logical (according to the model) process of reasoning. The subject then selects from the available list of conclusions one that matches the one that he or she has generated (or, in the case of Johnson-Laird & Steedman, 1978, he or she simply writes down the conclusion generated).

None of these models can, in its present form, account for the results of the present study. The main sources of error permitted by these models are either faulty representation of the premises (conversion) or figural bias in the processing of representations. Our syllogisms were constructed such that all premises were legally convertible; the figures chosen were those least susceptible to figural bias and, in any case, were consistent across conditions. Finally, the results cannot be explained by atmospheric bias, either (cf. Begg & Denny, 1969; Woodworth & Sells, 1935), since all conclusions were equally favored by atmosphere.

Two of the models provide additional scope for the occurrence of reasoning errors; those of Guyote and Sternberg (1981) and Johnson-Laird and Steedman (1978). Only Guyote and Sternberg have made an attempt to explain the effects of problem content on reasoning. One of their experiments compared reasoning with factual (believable) and counterfactual (unbelievable) content, but they do not discuss possible interactions with validity. They do say that content affected reasoning and that the subsequent parameter estimations for their model suggest that “subjects store and manipulate factual information with greater ease than they do other kinds of information (Guyote & Sternberg, 1981, p. 499). This implies that subjects should reason more logically with believable than with unbelievable content. In fact, the interaction observed in the present experiments was the opposite of this: Subjects were more sensitive to logical validity on unbelievable problems.

We do not believe that our results can be explained on the assumption that all reasoning proceeds from the representation of the premises toward a conclusion. It appears that subjects not only check the validity of the conclusion (by reference to the premises) but are also influenced by a separate, direct assessment of its truth value. There are several ways in which the Belief by Validity interaction could arise. First, it may be that subjects accept uncritically a conclusion with which they agree but are more likely to check the logic if they do not agree with the conclusion. This is directly analogous to the finding of Lord, Lepper, and Ross (1979) that people will accept at face value the evidence of research studies whose conclusions agree with their prior beliefs, but they will criticize the design and methodology of those with conflict.

There are however, other explanations of the interaction. In the valid-unbelievable condition, the conflict is that logic dictates that the conclusion must be accepted despite its unbelievable. However, in the invalid-believable condition, the conflicting role of logic is less strong. Logically, the invalid conclusions do not necessarily follow from the premises, but neither are they contradicted by them. Since the conclusion is not
inconsistent with the premises, subjects may feel justified in favoring belief. Dickstein (1980, 1981) has presented evidence that subjects may indeed have difficulty in understanding this aspect of the concept of logical necessity. It is possible that the somewhat weaker Belief by Logic interaction observed in Experiment 3 was due to the modification of the instructions that emphasized that "yes" answers should be given if only if the conclusions followed logically from the premises. This second explanation differs from the first in assuming that subjects always evaluate both the validity and the believability of each conclusion, but they respond differently to the two types of conflict.

A third explanation is that the conflict arising from unbelievable conclusions is less strong than that arising from believable conclusions, with our particular materials. Inspection of Table 1 reveals that while "true" statements are rated very close to the top of the scale, "false" items are rated, on average, only 1 point lower than the midpoint of the scale. Thus, in the valid-unbelievable condition, the bias to reject the conclusion on the basis of belief may be less strong than is the bias to accept, in accordance with belief, in the invalid-believable condition.

Unlike previous studies, we also have verbal protocol data to consider. The combined data of the three experiments are shown in two different ways in Table 8. Table 8a shows the probability of giving a particular explanation as a function of the response made. This is the appropriate way to look at the data if one assumes that they are rationalizations. In addition to the interaction that has been noted for the valid-unbelievable condition, a trend in the other conflict emerges on these pooled data. It seems that there is a tendency to give more references to irrelevant information when accepting invalid but believable conclusions. This suggests that subjects do perceive a conflict between logic and belief in this condition also, although less markedly than for valid-unbelievable problems.

On the other hand, if one supposes that the protocols do reflect the basis on which subjects were reasoning, then it is more appropriate to look at the likelihood of responses, given the protocol scores (see Table 8b). An interesting picture emerges here. It seems that the Logic by Belief interaction is present for problems in which subjects refer to the premises or do not make irrelevant references, but it is absent when the premises are not cited or when irrelevant information is mentioned. The latter problems show almost pure belief bias, with little effect at all of validity.

This might seem to suggest that there are two kinds of subjects: some resting their conclusions on the premises, and others not. However, although there is generally a negative relation between scores on the two criteria on any particular subject's response to a given problem, most subjects score positively on both criteria somewhere on their problems. It is still of interest to know whether subjects respond in a consistent manner to the conflict created by the valid-unbelievable condition. This can be examined by comparing the subjects' responses to the Figure 2 and 3 problems of this type in Experiment 3. It turns out that the response rates on the two tasks are quite independent: of the 16 accepting the Figure 2 conclusion, 9 accepted a Figure 3; of the 16 who did not accept a Figure 2, 8 accepted a Figure 3.

The above analysis supports the idea of a within-subjects conflict, as opposed to individual differences in strategies, which accords with Evans' (1977b) discussion of the Wason selection task. Indeed, the whole pattern of results is consistent with the theory of reasoning put forward by Evans (1982) and previously applied to conditional reasoning problems. That is, response probabilities reflect competing logical and nonlogical processes. However, we still have two areas of uncertainty in the interpretation of the results: (1) Which explanation of the Belief by Logic interaction is correct? and (2) to what extent do the protocols reflect rationalizations, and to what extent the actual basis of subjects reasoning? An additional treatment of the protocols, shown in Table 9, helps to provide answers to both these questions.

This analysis was concerned with the order of mention of the premises and conclusion. The analysis was confined to "thinking-aloud" protocols only (Experiment 2, Group 3, and Experiment 3 combined), since it was thought that order of items in retrospective protocols need not reflect the actual order in which the subjects did things. There were three main classifica-

<table>
<thead>
<tr>
<th>Table 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Positive Protocol Scores (Mentions) as a Function of Response Given and Percentage of Yes Responses as a Function of Protocol Scores</td>
</tr>
<tr>
<td>(a) Percentage of Positive Protocol Scores</td>
</tr>
<tr>
<td>Both Premises</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Valid-Believable</td>
</tr>
<tr>
<td>Valid-Unbelievable</td>
</tr>
<tr>
<td>Invalid-Believable</td>
</tr>
<tr>
<td>Invalid-Unbelievable</td>
</tr>
<tr>
<td>(b) Percentage of Yes Responses</td>
</tr>
<tr>
<td>Both Premises</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>Valid-Believable</td>
</tr>
<tr>
<td>Valid-Unbelievable</td>
</tr>
<tr>
<td>Invalid-Believable</td>
</tr>
<tr>
<td>Invalid-Unbelievable</td>
</tr>
</tbody>
</table>

Note - M = mentioned; NM = not mentioned. Data are combined for the three experiments; average Figure 2 and Figure 3 responses were used for Experiment 3 (n = 104).
tions: Conclusions only (C)—These protocols refer to the conclusion but do not mention either premise. They may or may not include references to irrelevant information. Conclusion to premises (CP)—These protocols included reference to at least one premise after mention of the conclusion. Premises to conclusions (PC)—These protocols included mention of at least one premise followed by mention of the conclusion. (In some cases, CP and PC protocols also included irrelevant information.)

Table 9a shows that the great majority of protocols were classifiable in one of these three ways and that the distribution of classifications over the four problem types was quite similar. Table 9b shows the percentage of subjects favoring belief (saying "yes" on believable or "no" on unbelievable problems) as a function of protocol classification. Several features of this table suggest that the protocol analysis is diagnostic of the amount of logical reasoning subjects are doing. The PC protocols are associated with least belief bias in the two conflict conditions, suggesting that they do reflect more logical (premise-to-conclusion) reasoning. Note also that almost all logical errors in the invalid-unbelievable condition are associated with PC protocols. This confirms the interpretation given in discussion of Experiment 2 that such errors arise from subjects who ignore belief, reason from the premises, but make a logical error in doing so.

The C protocols, we suggest, are those of subjects who focus their attention on the conclusion and thus give the highest rates of belief bias. Even here, however, there are two sources of evidence that the premises, although not mentioned, have some influence on the subjects. One is the visible fact in Table 9b that their rate of favoring belief is higher if the response is also logically correct. The second is the fact that the mention of irrelevant information by such subjects (as inferred from Table 8) is greater when the logic contradicts the belief-biased response. Thus C protocols are primarily associated with belief-biased, rationalizing subjects.

The CP protocols reflect subjects who focus on the conclusion but also go on to consider the premises. They show intermediate rates of belief bias, presumably because in some cases the premises are seen to contra-

dict the belief, but they still do less well than those who attempt to reason from premises to conclusion.

This interpretation of Table 9b suggests that there is no singular answer to the problem posed by Table 8: The protocols partially reflect the basis of subjects' reasoning and partially reflect rationalization. We now ask what Table 9 can contribute to our understanding of the Belief by Logic interaction. The frequency of PC protocols is unaffected by the believability of the conclusion in Table 9a (26% with believable and 26% with unbelievable conclusions). This is not surprising if the PC protocols indicate subjects engaged in premise-to-conclusion reasoning. However, our first interpretation of the interaction does predict a shift between C and CP protocols for subjects who focus initially on the conclusion. Specifically, it was suggested that such subjects are more likely to go on to consider the premises if the conclusion is unbelievable than if it is believable. Some shift in this direction is actually observed. There were 43% C protocols for believable problems, dropping to 38% on unbelievable; there were 27% CP protocols on believable, rising to 31% on unbelievable.

The general picture of Table 9 does not, however, support this interpretation of the interaction, for two reasons. First, the observed shift between C and CP protocols is too small to account for the large interaction in the response frequencies. Second, both CP and PC protocols are associated with substantially more belief bias in the invalid-believable condition than in the valid-unbelievable condition. This must mean that subjects who take account of both logic and belief experience more competition from belief in the former condition than in the latter. Thus, either our second or our third interpretation of the interaction is to be preferred.

Whether it is due to weaker logic on invalid than on valid problems or to stronger belief on believable than on unbelievable problems cannot be distinguished in the present experiments.

Finally, the analysis shown in Table 9 is helpful in resolving a problem raised by a reviewer of this paper, namely, that believability of the conclusion is inevitably confounded with believability of the premises. By the laws of logic, any valid argument with a false conclusion...
must have either a false premise or premises with incompatible suppositions. Such premises might strike the subject as anomalous and thus inhibit reasoning. However, the data of Table 9 strongly suggest that this is not the main cause of belief bias on valid-unbelievable problems. Table 9a shows that most subjects use a conclusion-centered strategy on these problems, and Table 9b shows that the majority (70%) of those who do reason from premises to conclusion correctly accept the inference.

In conclusion, we hope that we have shown that the introduction of protocol analysis has proved to be a more productive way of differentiating and understanding the processes underlying the belief-bias effect. The picture that finally emerges is that logic and belief conflict throughout, but they do so at different levels. When subjects focus primarily on the conclusion, belief biases are maximal and logical effects are minimal. Such subjects often rationalize their responses by referring to irrelevant information. These are the clearest examples of Henle’s (1962) “failure to accept the logical task,” but they still show a small influence of logic. On about 25% of occasions, though, a genuine premise-to-conclusion inference is attempted, with much higher logical success. It is most important to note, however, that even in cases in which the logical task is accepted, substantial (although lesser) effects of belief bias are still observed. These findings not only provide a challenge for existing models of syllogistic reasoning but also raise broader questions about people’s rational competence to generate and assess logical arguments in real life, whenever they have clear prior beliefs about the subject under discussion.

REFERENCES


DICKSTEIN, L. S. The effect of figure on syllogistic reasoning. *Cognition*, 1978, 6, 76-83. (a)

DICKSTEIN, L. S. Error processes in syllogistic reasoning. *Memory & Cognition*, 1978, 6, 537-543. (b)


EVANS, J. St. B. T. Toward a statistical theory of reasoning.


(Received for publication July 6, 1982; revision accepted January 31, 1983.)